

WO0008157

Publication Title:

HUMAN ANION TRANSPORTER GENES ATNOV

Abstract:

126b Abstract of WO0008157

Methods for isolating ATnov genes are provided. The ATnov nucleic acid compositions find use in identifying homologous or related proteins and the DNA sequences encoding such proteins; in producing compositions that modulate the expression or function of the protein; and in studying associated physiological pathways. In addition, modulation of the gene activity *in vivo* is used for prophylactic and therapeutic purposes, such as identification of cell type based on expression, and the like. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, 5/10, C12Q 1/68, A01K 67/00, C12N 15/00, C07K 16/28, 14/705		A2	(11) International Publication Number: WO 00/08157 (43) International Publication Date: 17 February 2000 (17.02.00)
(21) International Application Number: PCT/US99/17823 (22) International Filing Date: 6 August 1999 (06.08.99)		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 60/095,835 7 August 1998 (07.08.98) US		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(71) Applicant (for all designated States except US): AXYS PHARMACEUTICALS, INC. [US/US]; 180 Kimball Way, South San Francisco, CA 94080 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): LAUBERT, Boris [US/US]; 4550 Bancroft Street #4, San Diego, CA 92116 (US). CARDOSO, Gizela [US/US]; 721 North Main Street, Brockton, Maine 02401 (US). HU, Ping [US/US]; 5807 Folkstone Road, Bethesda, MD 20817 (US). MILLER, Andrew, P. [US/US]; 3271 Countryside Drive, San Mateo, CA 94403 (US). BUCKLER, Alan, J. [US/US]; 2315 Lagoon View Drive, Cardiff, CA 92007 (US).			
(74) Agent: SHERWOOD, Pamela, J.; Bozicevic, Field & Francis LLP, Suite 200, 285 Hamilton Avenue, Palo Alto, CA 94301 (US).			

(54) Title: HUMAN ANION TRANSPORTER GENES ATNOV

(57) Abstract

Methods for isolating ATnov genes are provided. The ATnov nucleic acid compositions find use in identifying homologous or related proteins and the DNA sequences encoding such proteins; in producing compositions that modulate the expression or function of the protein; and in studying associated physiological pathways. In addition, modulation of the gene activity in vivo is used for prophylactic and therapeutic purposes, such as identification of cell type based on expression, and the like.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN ANION TRANSPORTER GENES

INTRODUCTION

Background

5 Endo- and xenobiotics are typically cleared from mammals via the liver, the primary site of drug metabolizing enzymes. Charged compounds, either endogenously or exogenously derived, are taken up by hepatocytes across the basolateral membrane, appropriately metabolized by the liver enzymes, trafficked through the cell, and then excreted across the canalicular membrane into the bile. These four steps are important in
10 determining a patient's response to pharmaceutical agents.

Generation of bile flow is a regulated, ATP-dependent process and depends on the coordinated action of a number of transporter proteins in the sinusoidal and canalicular domains of the hepatocyte. Dysfunction of any of these proteins leads to retention of substrates, with conjugated hyperbilirubinemia or cholestasis as a result. In recent years
15 many of the transport proteins involved in bile formation have been identified, cloned, and functionally characterized. The hepatocyte sinusoidal membrane contains transport proteins for the hepatic uptake of organic anions and cations and for the uptake of bile acids.

20 The Na⁺-independent organic anion transporter, OATP, resides on the basolateral surface of hepatocytes and mediates the uptake of a large number of amphipathic substrates, such as bromosulfophalein, bile acids, estrogen conjugates, neutral steroids, organic cations, cardiac glycosides, and peptidomimetic drugs. The human organic anion transporter, OATP, is expressed in multiple tissues, including brain, lung, liver, kidney, and testes, while a rat homolog of OATP is expressed only in liver and kidney (Bergwerk et al.
25 (1996) *Am. J. Physiol.* 271: G231-G238). A prostaglandin transporter, hPGT, which shares significant homology with these organic anion transporters, is abundantly expressed (Lu et al. (1996) *J. Clin. Invest.* 98: 1142-1149).

30 Variations in transporter sequences may alter the kinetic properties of the protein. For example, inefficient clearance of substrates would result in an increased biological half-life, where drugs have an increased half-life and drug levels approach or reach toxic thresholds. Alternatively, over-efficient clearance of substrates could reduce the biological effectiveness of a drug. The identification of novel genes within these pathways provides additional targets for pharmacogenetic analysis, as well as a more thorough understanding of the biological process of drug clearance.

Relevant Literature

The molecular and functional characterization of an organic anion transporting polypeptide cloned from human liver, OATP, is described by Kullak-Ublick et al. (1995)

5 Gastroenterology 109:1274-1282. Other cloned transporter genes are described by Noe et al. (1997) Proc. Natl. Acad. Sci. 94:10346-10350; and Jacquemin et al. (1994) Proc. Natl. Acad. Sci. 91:133-137.

The role of organic cation transporters in intestine, kidney, liver, and brain is reviewed by Koepsell (1998) Annu Rev Physiol 60:243-266. Canicular multispecific organic anion transporter and the disposal of endo- and xenobiotics is reviewed by Elferink and Jansen (1994) Pharmac. Ther. 64:77-97.

10 Public EST sequences having sequence similarity with ATnov nucleic acids include: Genbank accessions nos. N49902 (ATnov2); N50005 (ATnov2); H62927 (ATnov3); H62893 (ATnov3); R29414 (ATnov3); AA382692 (ATnov3); T73863 (ATnov3); T74263 (ATnov3);
15 T55488 (ATnov3).

SUMMARY OF THE INVENTION

Isolated nucleotide compositions and sequences are provided for ATnov genes. The ATnov nucleic acid compositions find use in identifying homologous or related genes; in
20 producing compositions that modulate the expression or function of its encoded proteins; for gene therapy; mapping functional regions of the proteins; and in studying associated physiological pathways. In addition, modulation of the gene activity in vivo is used for prophylactic and therapeutic purposes, such as treatment of anion transporter defects, identification of cell type based on expression, and the like.

25 DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Nucleic acid compositions encoding ATnov anion transporters are provided. They are used in identifying homologous or related genes; in producing compositions that modulate the expression or function of the encoded proteins; for gene therapy; mapping functional regions of the proteins; and in studying associated physiological pathways. The
30 ATnov gene products are members of the anion transporter gene family, and have high degrees of homology at the amino acid level with known anion transporters.

CHARACTERIZATION OF ATNOV

The sequence data predict that the provided ATnov genes encode anion transporters. Characterization of organic ion transport across the cell membrane, in terms of substrates, binding and transport kinetics, is an important aspect of ATnov biology. A 5 substrate, as used herein, is a chemical entity that is transported by an ATnov polypeptide, usually under normal physiological conditions. Substrates can be either endogenous substrates, i.e. substrates normally found within the natural environment, such as bile salts, or exogenous, i.e. substrates that are not normally found within the natural environment.

Substrate screening assays are used to determine the kinetics of a ATnov protein or 10 peptide fragment on a substrate. Many suitable assays are known in the art, including the use of primary or cultured cells, genetically modified cells (e.g., where DNA encoding the ATnov polymorphism to be studied is introduced into the cell within an artificial construct), cell-free systems, e.g. recombinantly produced enzymes in a suitable buffer, or in animals, including human clinical trials (see, e.g. (1995) Burchell et al. Life Sci. 57:1819-1831, 15 specifically incorporated herein by reference). Where genetically modified cells are used, since most cell lines do not express ATnov activity (liver cells lines being the exception), introduction of artificial construct for expression of the ATnov polymorphism into many human and non-human cell lines does not require additional modification of the host to inactivate endogenous ATnov expression/activity. Clinical trials may monitor serum, urine, 20 etc. levels of the substrate or its metabolite(s).

Full length ion transporter cDNAs may be combined with proper vectors to form 25 expression constructs of each individual transporter. Functional analyses of expressed transporters can be performed in heterologous systems, or by expression in mammalian cell lines. For expression analyses in heterologous systems such as *Xenopus* oocytes, synthetic mRNA is made through in vitro transcription of each transporter construct. mRNA is then injected into prepared oocytes and the cells allowed to express the transporter for several days. Candidate substrates may be labeled to provide a means of following movement across the membrane. Similarly, the requirements of a transporter for ATP, Na⁺, etc. may be assessed. For an example of these techniques, see Kullak-Ublick et al. (1997) 30 Gastroenterology 113(4):1295-1305.

Heterologous or mammalian cell lines expressing the novel transporters can be used to characterize small molecules and drugs that interact with the transporter. The same

experiments can be used to assay for novel compounds that interact with the expressed transporters.

ATNOV NUCLEIC ACID COMPOSITIONS

5 As used herein, the term "ATnov" is generically used to refer to any one of the provided nucleotide sequences as set forth in the SEQLIST. Of particular interest are the sequences, including polymorphisms, of ATnov3.1 and ATnov3.2. These sequences are provided as SEQ ID NO:3 (ATnov3.1), SEQ ID NO:5 (ATnov3.1), SEQ ID NO:7 (ATnov3.2) and SEQ ID NO:9 (ATnov3.2). The encoded polypeptides are provided as SEQ ID NO:4, 10 6, 8 and 10, respectively. The polymorphic variants are set forth in the sequences listings. These include a G or A polymorphism at nucleotide 487, resulting in an amino acid change of asp to asn. There is a polymorphism of C or T at nucleotide 670, which is silent with respect to the encoded polypeptide. A frameshift variant is found in the poly T stretch between positions 1705 and 1710, where the sequence contains either 5T or 6T. The 5T 15 polymorphism results in a truncated polypeptide product of 542 amino acids (SEQ ID NO:4 and SEQ ID NO:8), while the 6T polymorphism encodes the full-length protein of 591 amino acids.

Also of interest are the genetic sequences of SEQ ID NO:1 (ATnov1) and SEQ ID NO:2 (ATnov2).

20 Where a specific ATnov sequence is intended, the numerical designation will be added. Nucleic acids encoding ATnov anion transporters may be cDNA or genomic DNA or a fragment thereof. The term "ATnov gene" shall be intended to mean the open reading frame encoding any of the provided ATnov polypeptides, introns, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 25 kb beyond the coding region, but possibly further in either direction. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

Novel nucleic acid compositions of the invention of particular interest comprise a sequence set forth in SEQ ID NO:1, 2, 3, 5, 7, 9 or an identifying sequence thereof. An 30 "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a nucleic acid sequence, e.g., exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt.

Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from SEQ ID NO:1, 2, 3, 5, 7, 9.

The nucleic acids of the invention also include nucleic acids having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M NaCl/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM NaCl/0.9 mM sodium citrate).

Hybridization methods and conditions are well known in the art, see U.S. Patent No. 5,707,829. Nucleic acids that are substantially identical to the provided nucleic acid sequences, e.g. allelic variants, genetically altered versions of the gene, etc., bind to the provided nucleic acid sequences (SEQ ID NO:1, 2, 3, 5, 7, 9) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species.

Preferably, hybridization is performed using at least 15 contiguous nucleotides of SEQ ID NO:1, 2, 3, 5, 7, 9. The probe will preferentially hybridize with a nucleic acid or mRNA comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids of the biological material that uniquely hybridize to the selected probe.

Probes of more than 15 nucleotides can be used, e.g. probes of from about 18 nucleotides to not more than about 100 nucleotides, but 15 nucleotides generally represents sufficient sequence for unique identification.

The nucleic acids of the invention also include naturally occurring variants of the nucleotide sequences, e.g. degenerate variants, allelic variants, etc. Variants of the nucleic acids of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the nucleic acids of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair mismatches relative to the selected nucleic acid probe. In general, allelic variants contain 5-25% base pair mismatches, and can contain as little as even 2-5%, or 1-2% base pair mismatches, as well as a single base-pair mismatch.

The invention also encompasses homologs corresponding to the nucleic acids of SEQ ID NO:1, 2, 3, 5, 7, 9, where the source of homologous genes can be any related

species within the same genus or group. Within a group, homologs have substantial sequence similarity, e.g. at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al., *J. Mol. Biol.* (1990) 215:403-10.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90% or more as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

ATnov polymorphic sequences. It has been found that specific sites in the ATnov gene sequence are polymorphic, i.e. within a population, more than one nucleotide (G, A, T, C) is found at a specific position. Polymorphisms may provide functional differences in the genetic sequence, through changes in the encoded polypeptide, changes in mRNA stability, binding of transcriptional and translation factors to the DNA or RNA, and the like. The polymorphisms are also used as single nucleotide polymorphisms to detect association with, or genetic linkage to phenotypic variation in activity and expression of ATnov.

SNPs are generally biallelic systems, that is, there are two alleles that an individual may have for any particular marker. SNPs, found approximately every kilobase, offer the potential for generating very high density genetic maps, which will be extremely useful for developing haplotyping systems for genes or regions of interest, and because of the nature of SNPs, they may in fact be the polymorphisms associated with the disease phenotypes under study. The low mutation rate of SNPs also makes them excellent markers for studying complex genetic traits.

Single nucleotide polymorphisms are provided in the ATnov3 sequence listing. The provided sequences also encompass the complementary sequence corresponding to any of the provided polymorphisms.

In order to provide an unambiguous identification of the specific site of a polymorphism, sequences flanking the polymorphic site are included in a probe for the region. It will be understood that there is no special significance to the length of non-polymorphic flanking sequence that is included, except to aid in positioning the polymorphism in the genomic sequence.

For screening purposes, hybridization probes of the polymorphic sequences may be used where both forms are present, either in separate reactions, spatially separated on a solid phase matrix, or labeled such that they can be distinguished from each other. Assays may utilize nucleic acids that hybridize to one or more of the described polymorphisms.

An array may include all or a subset of the ATnov3 polymorphisms. One or both polymorphic forms may be present in the array. Usually such an array will include at least 2 different polymorphic sequences, i.e. polymorphisms located at unique positions within the locus, and may include as many all of the provided polymorphisms. Arrays of interest may further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for pharmacogenetic screening. The oligonucleotide sequence on the array will usually be at least about 12 nt in length, may be the length of the provided polymorphic sequences, or may extend into the flanking regions to generate fragments of 100 to 200 nt in length. For examples of arrays, see Ramsay (1998) Nat. Biotech. 16:40-44; Hacia et al. (1996) Nature Genetics 14:441-447; Lockhart et al. (1996) Nature Biotechnol. 14:1675-1680; and De Risi et al. (1996) Nature Genetics 14:457-460.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein. The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the

introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end 5 of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for expression.

10 The nucleic acid compositions of the subject invention can encode all or a part of the subject polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, etc. Isolated nucleic acids and nucleic acid fragments of the invention comprise at least about 15 up to about 100 contiguous nucleotides, or up to the complete sequence provided in SEQ ID NO:1, 2, 3, 5, 15 7 or 9. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more.

20 Probes specific to the nucleic acids of the invention can be generated using the nucleic acid sequences disclosed in SEQ ID NO:1, 2, 3, 5, 7 or 9 and the fragments as described above. The probes can be synthesized chemically or can be generated from longer nucleic acids using restriction enzymes. The probes can be labeled, for example, with 25 a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a nucleic acid of one of SEQ ID NO:1, 2, 3, 5, 7 or 9. More preferably, probes are designed based on a contiguous sequence of one of the subject nucleic acids that remain unmasked following application of a masking program for masking low complexity (e.g., XBLAST) to the sequence., i.e. one would select an unmasked region, as indicated by the nucleic acids outside the poly-n stretches of the masked sequence produced by the masking program.

30 The nucleic acids of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the nucleic acids, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", e.g., flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The nucleic acids of the invention can be provided as a linear molecule or within a circular molecule. They can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. They can be regulated by their own or by other regulatory sequences, as is known in the art. The nucleic acids of the invention can

5 be introduced into suitable host cells using a variety of techniques which are available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

10 The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological samples (e.g., extracts of cells) to generate additional copies of the nucleic acids, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example,

15 determine the presence or absence of the nucleic acid sequences as shown in SEQ ID NO:1, 2, 3, 5, 7 or 9 or variants thereof in a sample.

The sequence of the 5' flanking region may be utilized for promoter elements, including enhancer binding sites, that provide for developmental regulation in tissues where ATnov genes are expressed. The tissue specific expression is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter regions are useful for determining natural variations in expression, particularly those that may be associated with disease.

Alternatively, mutations may be introduced into the promoter regions to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, e.g. sequence similarity to known binding motifs, gel retardation studies, etc. For examples, see Blackwell et al. (1995) Mol Med 1: 194-205; Mortlock et al. (1996) Genome Res. 6: 327-33; and Joulin and Richard-Foy (1995) Eur J Biochem 232: 620-626.

30 The regulatory sequences may be used to identify cis acting sequences required for transcriptional or translational regulation of ATnov expression, especially in different tissues or stages of development, and to identify cis acting sequences and trans acting factors that regulate or mediate ATnov expression. Such transcription or translational control regions

may be operably linked to a ATnov gene in order to promote expression of wild type or altered ATnov or other proteins of interest in cultured cells, or in embryonic, fetal or adult tissues, and for gene therapy.

Double or single stranded fragments may be obtained of the DNA sequence by 5 chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, etc. For the most part, DNA fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and may be at least about 50 nt. Such small DNA fragments are useful as primers for PCR, hybridization screening probes, etc. Larger DNA fragments, i.e. greater than 100 nt are useful for production of the encoded 10 polypeptide. For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an 15 amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

The DNA may also be used to identify expression of the gene in a biological specimen. The manner in which one probes cells for the presence of particular nucleotide 20 sequences, as genomic DNA or RNA, is well established in the literature and does not require elaboration here. DNA or mRNA is isolated from a cell sample. The mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject 25 DNA sequences. Alternatively, the mRNA sample is separated by gel electrophoresis, transferred to a suitable support, e.g. nitrocellulose, nylon, etc., and then probed with a fragment of the subject DNA as a probe. Other techniques, such as oligonucleotide ligation assays, in situ hybridizations, and hybridization to DNA probes arrayed on a solid chip may also find use. Detection of mRNA hybridizing to the subject sequence is indicative of ATnov gene expression in the sample.

30 The sequence of an ATnov gene, including flanking promoter regions and coding regions, may be mutated in various ways known in the art to generate targeted changes in promoter strength, sequence of the encoded protein, etc. The DNA sequence or protein product of such a mutation will usually be substantially similar to the sequences provided

herein, i.e. will differ by at least one nucleotide or amino acid, respectively, and may differ by at least two but not more than about ten nucleotides or amino acids. The sequence changes may be substitutions, insertions or deletions. Deletions may further include larger changes, such as deletions of a domain or exon. Other modifications of interest include 5 epitope tagging, e.g. with the FLAG system, HA, etc. For studies of subcellular localization, fusion proteins with green fluorescent proteins (GFP) may be used.

Techniques for in vitro mutagenesis of cloned genes are known. Examples of protocols for site specific mutagenesis may be found in Gustin et al., *Biotechniques* 14:22 (1993); Barany, *Gene* 37:111-23 (1985); Colicelli et al., *Mol Gen Genet* 199:537-9 (1985); 10 and Prentki et al., *Gene* 29:303-13 (1984). Methods for site specific mutagenesis can be found in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, CSH Press 1989, pp. 15.3-15.108; Weiner et al., *Gene* 126:35-41 (1993); Sayers et al., *Biotechniques* 13:592-6 (1992); Jones and Winstorfer, *Biotechniques* 12:528-30 (1992); Barton et al., *Nucleic Acids Res* 18:7349-55 (1990); Marotti and Tomich, *Gene Anal Tech* 6:67-70 (1989); and Zhu, *Anal 15 Biochem* 177:120-4 (1989). Such mutated genes may be used to study structure-function relationships of ATnov polypeptides, or to alter properties of the protein that affect its function or regulation.

Genetic polymorphisms, either naturally occurring or introduced as described above, are useful in screening for altered transport or metabolism of ATnov substrates. For 20 example, variant alleles may affect the pharmacokinetic parameters of substrates. A drug's volume of distribution, clearance, and the derived parameter, half-life, are particularly important, as they determine the degree of fluctuation between a maximum and minimum plasma concentration during a dosage interval, the magnitude of steady state concentration and the time to reach steady state plasma concentration upon chronic dosing. Parameters 25 derived from in vivo drug administration are useful in determining the clinical effect of a particular ATnov genotype.

ATNOV POLYPEPTIDES

The subject gene may be employed for producing all or portions of ATnov 30 polypeptides. Fragments of interest include the glycosylation sites, transmembrane domains, ATP binding regions, the substrate binding sites, etc. Such domains will usually include at least about 20 amino acids of the provided sequence, more usually at least about 50 amino acids, and may include 100 amino acids or more, up to the complete domain.

Binding contacts may be comprised of non-contiguous sequences, which are brought into proximity by the tertiary structure of the protein. The sequence of such fragments may be modified through manipulation of the coding sequence, as described above. Truncations may be performed at the carboxy or amino terminus of the fragment, e.g. to determine the 5 minimum sequence required for biological activity.

A subset of the provided nucleic acid polymorphisms in ATnov3 confer a change in the corresponding amino acid sequence, as previously described. Using the amino acid sequence provided in SEQ ID NO:3 as a reference, the amino acid polymorphisms of the invention include asn \square asp, pos. 130; and a frameshift at position 537 resulting in a truncated 10 protein of 542 amino acids. Polypeptides comprising at least one of the provided polymorphisms (ATnov3^v polypeptides) are of interest. The term "ATnov3^v polypeptides" as used herein includes complete ATnov protein forms, e.g. such splicing variants as known in the art, and fragments thereof, which fragments may comprise short polypeptides, epitopes, functional domains; binding sites; etc.; and including fusions of the subject polypeptides to 15 other proteins or parts thereof. Polypeptides will usually be at least about 8 amino acids in length, more usually at least about 12 amino acids in length, and may be 20 amino acids or longer, up to substantially the complete protein.

For expression, an expression cassette may be employed. The expression vector will provide a transcriptional and translational initiation region, which may be inducible or 20 constitutive, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. These control regions may be native to an ATnov gene, or may be derived from exogenous 25 sources.

The peptide may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production 30 of the protein, a unicellular organism, such as *E. coli*, *B. subtilis*, *S. cerevisiae*, insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g. COS 7 cells, may be used as the expression host cells. In some situations, it is desirable to express the ATnov gene in eukaryotic cells, where the ATnov protein will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory. Peptides that are subsets of the complete ATnov sequence may be used to identify and investigate parts of the protein important for function, or to raise antibodies directed against these regions.

With the availability of the protein or fragments thereof in large amounts, by employing an expression host, the protein may be isolated and purified in accordance with conventional ways. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or 5 other purification technique. The purified protein will generally be at least about 80% pure, preferably at least about 90% pure, and may be up to and including 100% pure. Pure is intended to mean free of other proteins, as well as cellular debris.

The expressed ATnov polypeptides are useful for the production of antibodies, where short fragments provide for antibodies specific for the particular polypeptide, and larger 10 fragments or the entire protein allow for the production of antibodies over the surface of the polypeptide. Antibodies may be raised to the wild-type or variant forms of ATnov. Antibodies may be raised to isolated peptides corresponding to these domains, or to the native protein.

Antibodies are prepared in accordance with conventional ways, where the expressed 15 polypeptide or protein is used as an immunogen, by itself or conjugated to known immunogenic carriers, e.g. KLH, pre-S HBsAg, other viral or eukaryotic proteins, or the like. Various adjuvants may be employed, with a series of injections, as appropriate. For 20 monoclonal antibodies, after one or more booster injections, the spleen is isolated, the lymphocytes immortalized by cell fusion, and then screened for high affinity antibody binding. The immortalized cells, i.e. hybridomas, producing the desired antibodies may then be 25 expanded. For further description, see Monoclonal Antibodies: A Laboratory Manual, Harlow and Lane eds., Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1988. If desired, the mRNA encoding the heavy and light chains may be isolated and mutagenized by cloning in *E. coli*, and the heavy and light chains mixed to further enhance the affinity of the antibody. Alternatives to *in vivo* immunization as a method of raising antibodies include binding to phage "display" libraries, usually in conjunction with *in vitro* affinity maturation.

ATNOV GENOTYPING

The subject nucleic acid and/or polypeptide compositions may be used in genotyping 30 and to screen for the presence of polymorphisms in the sequence, or variation in the expression of the subject genes. Genotyping may be performed to determine whether a particular polymorphisms is associated with a disease state or genetic predisposition to a disease state, particularly diseases associated with liver disorders.

Genotyping may also be performed for pharmacogenetic analysis to assess the association between an individual's genotype and that individual's ability to react to a therapeutic agent. Differences in substrate transport to relevant cells can lead to toxicity or therapeutic failure. Relationships between polymorphisms in transporter expression or specificity can be used to optimize therapeutic dose administration.

ATnov genotyping is performed by DNA or RNA sequence and/or hybridization analysis of any convenient sample from a patient, e.g. biopsy material, blood sample, scrapings from cheek, etc. A nucleic acid sample from an individual is analyzed for the presence of polymorphisms in ATnov, particularly those that affect the activity, responsiveness or expression of ATnov. Specific sequences of interest include any polymorphism that leads to changes in basal expression in one or more tissues, to changes in the modulation of ATnov expression, or alterations in ATnov specificity and/or activity.

The effect of a polymorphism in ATnov gene sequence on the response to a particular agent may be determined by in vitro or in vivo assays. Such assays may include monitoring during clinical trials, testing on genetically defined cell lines, etc. The response of an individual to the agent can then be predicted by determining the ATnov genotype with respect to the polymorphism. Where there is a differential distribution of a polymorphism by racial background, guidelines for drug administration can be generally tailored to a particular ethnic group.

Biochemical studies may be performed to determine whether a sequence polymorphism in a ATnov coding region or control regions is associated with disease. Disease associated polymorphisms may include deletion or truncation of the gene, mutations that alter expression level, that affect the specificity or transport kinetics of the transporter, etc.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki et al. (1985) Science 239:487, and a review of current techniques may be found in Sambrook et al. Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2-14.33. Amplification may be used to determine whether a polymorphism is present, by using a primer that is

specific for the polymorphism. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Delahunty et al. (1996) *Am. J. Hum. Genet.* 58:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels

5 include fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g. ^{32}P , ^{35}S , ^{3}H ; etc. The label may be a

10 two stage system, where the amplified DNA is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

15 The sample nucleic acid, e.g. amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, etc. The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilised on a solid support, as described

20 in U.S. 5,445,934, or in WO95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), mismatch cleavage detection, and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a

25 polymorphism creates or destroys a recognition site for a restriction endonuclease (restriction fragment length polymorphism, RFLP), the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

30 In one embodiment of the invention, an array of oligonucleotides are provided, where discrete positions on the array are complementary to one or more of the provided polymorphic sequences, e.g. oligonucleotides of at least 12 nt, frequently 20 nt, or larger, and including the sequence flanking the polymorphic position. Such an array may comprise

a series of oligonucleotides, each of which can specifically hybridize to a different polymorphism. For examples of arrays, see Hacia et al. (1996) Nature Genetics 14:441-447; Lockhart et al. (1996) Nature Biotechnol. 14:1675-1680; and De Risi et al. (1996) Nature Genetics 14:457-460.

5 Screening for polymorphisms in ATnov may be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that may affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in ATnov proteins may be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to
10 be effective screening tools. The activity of the encoded ATnov protein as a anion transporter may be determined by comparison with the wild-type protein.

Antibodies specific for a ATnov may be used in staining or in immunoassays. Samples, as used herein, include biological fluids such as semen, blood, cerebrospinal fluid, tears, saliva, lymph, dialysis fluid and the like; organ or tissue culture derived fluids; and
15 fluids extracted from physiological tissues. Also included in the term are derivatives and fractions of such fluids. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

Diagnosis may be performed by a number of methods to determine the absence or presence or altered amounts of normal or abnormal ATnov polypeptides in patient cells. For
20 example, detection may utilize staining of cells or histological sections, performed in accordance with conventional methods. The antibodies of interest are added to the cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other labels for direct detection. Alternatively, a second
25 stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Alternatively, the secondary antibody conjugated to a fluorescent compound, e.g. fluorescein, rhodamine, Texas red, etc. Final detection uses a substrate that undergoes a color change in the
30 presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, etc.

MODULATION OF GENE EXPRESSION

The ATnov genes, gene fragments, or the encoded protein or protein fragments are useful in gene therapy to treat disorders associated with ATnov defects. Expression vectors may be used to introduce the ATnov gene into a cell. Such vectors generally have 5 convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, e.g. plasmid; retrovirus, e.g. lentivirus; adenovirus; and the like, where the vectors are able 10 to transiently or stably be maintained in the cells, usually for a period of at least about one day, more usually for a period of at least about several days to several weeks.

The gene or ATnov protein may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth et al. (1992) Anal 15 Biochem 205:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang et al. (1992) Nature 356:152-154), where gold microprojectiles are coated with the ATnov or DNA, then bombarded into skin cells.

Antisense molecules can be used to down-regulate expression of ATnov in cells. 20 The anti-sense reagent may be antisense oligonucleotides (ODN), particularly synthetic ODN having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, e.g. by reducing 25 the amount of mRNA available for translation, through activation of RNase H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences.

Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such 30 that an antisense strand is produced as an RNA molecule. Alternatively, the antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about

35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and selective inhibitors of gene expression (see Wagner et al. (1996) Nature Biotechnology 14:840-844).

5 A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an in vitro or animal model. A combination of sequences may also be used, where several regions of the mRNA sequence
10 are selected for antisense complementation.

Antisense oligonucleotides may be chemically synthesized by methods known in the art (see Wagner et al. (1993) *supra*. and Milligan et al., *supra*.) Preferred oligonucleotides are chemically modified from the native phosphodiester structure, in order to increase their intracellular stability and binding affinity. A number of such modifications have been
15 described in the literature, which alter the chemistry of the backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3 \square -O-5 \square -S-phosphorothioate, 3 \square -S-5 \square -O-phosphorothioate, 3 \square -CH₂-5 \square -O-phosphonate and 3 \square -NH-5 \square -O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage. Sugar modifications are also used to enhance stability and affinity. The \square -anomer of deoxyribose may be used, where the base is inverted with respect to the natural \square -anomer. The 2 \square -OH of the ribose sugar
20 may be altered to form 2 \square -O-methyl or 2 \square -O-allyl sugars, which provides resistance to degradation without comprising affinity. Modification of the heterocyclic bases must maintain
25 proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2 \square -deoxycytidine and 5-bromo-2 \square -deoxycytidine for deoxycytidine. 5- propynyl-2 \square -deoxyuridine and 5-propynyl-2 \square -deoxycytidine have been shown to increase affinity and
30 biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

As an alternative to anti-sense inhibitors, catalytic nucleic acid compounds, e.g. ribozymes, anti-sense conjugates, etc. may be used to inhibit gene expression. Ribozymes may be synthesized in vitro and administered to the patient, or may be encoded on an

expression vector, from which the ribozyme is synthesized in the targeted cell (for example, see International patent application WO 9523225, and Beigelman et al. (1995) Nucl. Acids Res 23:4434-42). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of anti-sense ODN with a metal complex, e.g. terpyridylCu(II), capable of mediating mRNA hydrolysis are described in Bashkin et al. (1995) Appl Biochem Biotechnol 54:43-56.

GENETICALLY ALTERED CELL OR ANIMAL MODELS FOR ATNOV FUNCTION

The subject nucleic acids can be used to generate transgenic animals or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination, where the normal ATnov locus is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like.

The modified cells or animals are useful in the study of ATnov function and regulation. For example, a series of small deletions and/or substitutions may be made in the ATnov gene to determine the role of different transmembrane domains, of ATP catalysis, etc. Of interest are the use of ATnov to construct transgenic animal models where expression of ATnov is specifically reduced or absent. Specific constructs of interest include anti-sense ATnov, which will block ATnov expression, expression of dominant negative ATnov mutations, etc. One may also provide for expression of the ATnov gene or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development.

DNA constructs for homologous recombination will comprise at least a portion of the ATnov gene with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown et al. (1990) Methods in Enzymology 185:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or embryonic cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of leukemia inhibiting factor (LIF). When ES or embryonic cells have been transformed, they may be used to

produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those 5 colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting offspring screened for the 10 construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be 15 maintained as allogeneic or congenic grafts or transplants, or in in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc.

TESTING OF ATNOV FUNCTION and RESPONSES

20 Anion transporters such as ATnov polypeptides are involved in multiple biologically important processes. Pharmacological agents designed to affect only specific transporter subtypes are of particular interest. The subject polypeptides may be used to test the specificity of novel compounds, and of analogs and derivatives of compounds known to be substrates, or to act on anion transporters.

25 Drug screening may be performed using an in vitro model, a genetically altered cell or animal, or purified ATnov protein. One can identify ligands or substrates that bind to, modulate or mimic the action of ATnov. Drug screening identifies agents that provide a replacement for ATnov function in abnormal cells. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be 30 used for this purpose, including monitoring cellular excitation and conductance, labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. The purified protein may also be used for determination of

three-dimensional crystal structure, which can be used for modeling intermolecular interactions.

The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of 5 ATnov polypeptide. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e. at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are 10 organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or 15 heterocyclic structures and/or aromatic or polycyclic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of 20 synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds 25 are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs.

Where the screening assay is a binding assay, one or more of the molecules may be 30 joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles, e.g. magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin etc. For the specific

binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc. that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions.

5 Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc. may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40°C. Incubation periods are selected for

10 optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host in a variety of ways, orally, topically, parenterally e.g. subcutaneously, intraperitoneally, by viral infection, intravascularly, etc.

15 Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%. The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for

20 oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents.

EXPERIMENTAL

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have
5 been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

10

Example 1

Three novel members of the OATP gene family, which are expressed in liver tissue, were cloned. These genes were isolated using trapped exons obtained from large-scale exon trapping of chromosome 12. The three anion transporters reported here are 70-80%
15 identical to each other over the predicted protein sequence, and are each 40% identical to the reported OATP protein sequence (Kullack-Ublick et al., (1995) Gastroenterology 109:1274-1282). The chromosomal location of these three anion transporters, along with the mapping of OATP, suggests this gene-family is clustered on 12p12.

20 Materials and Methods

cDNA Isolation. cDNA clones were isolated using the GeneTrapper system (Gibco-BRL). PCR primers within the trapped exons were used to detect which plasmid cDNA libraries contained the gene of interest. Oligonucleotide probes were designed: (SEQ ID NO:7) C12B_120: GGGGCTCTGATTGATACAACGTG ; (SEQ ID NO:8) C12C_151: ACTGTGGCACACGTGGTCATGTAGGACAT) and the process proceeded according to
25 the supplied protocol. cDNA clones were sequenced on an ABI 377 according to standard methods. The Primer Island Transposition kit was used according to the supplied protocol. Sequences were analyzed, edited, and assembled using the Sequencher software (Gene Codes).

30

Radiation Hybrid Mapping. RH mapping was achieved using the Stanford G3 panel DNAs (Research Genetics). DNA was aliquoted into 96-well trays, dried, and resuspended in PCR buffer prior to PCR amplification. 20 μ l PCR reactions with standard conditions, 2.5

mM MgCl₂, Taq Gold, and an annealing temperature of 60°C (for ATnov1 and 2) or 55°C (for ATnov3) were used to detect expression. The assays were done in duplicate and results were scored and map positions determined via the RH server at Stanford University <<http://www-shgc.stanford.edu/RH/G3index.html>>.

5

RT-PCR. RT-PCR was utilized to characterize the expression pattern of the novel anion transporters. This approach used RNA from 30 different tissues to generate first strand cDNA. Total RNA was purchased (Clontech, Invitrogen) and used to synthesize first strand cDNA using M-MLV reverse transcriptase and the supplied buffer (Gibco-BRL). The 10 20 µl reaction contained 5 µg total RNA, 100 ng of random primers, 10 mM DTT, 0.5 mM each dNTP, and an RNase inhibitor (Gibco-BRL). Identical reactions were set up without reverse transcriptase to control for DNA contamination in the RNA samples. The synthesis reaction proceeded for 1 hour at 37°C followed by 10 minutes at 95°C. These cDNAs, along 15 with control cDNA synthesis reactions without reverse transcriptase, were diluted 1:5 and 20 2 µl of each sample were arrayed into 96-well trays, dried, and resuspended in PCR buffer prior to PCR amplification. The cDNAs were tested with primers with defined expression patterns to verify the presence of amplifiable cDNA from each tissue. Gene-specific primers were used to amplify the cDNAs in 20 µl PCR reactions with standard conditions, 2.5 mM MgCl₂, Taq Gold, and an appropriate annealing temperature.

25 This approach provides for relatively high-throughput analysis of gene expression in a large set of tissues in a cost-efficient manner and provides qualitative analysis of gene expression only. Modifications can be employed, such as the use of internal control primers, limited cycling parameters, and dilution series to convert this to a quantitative experiment.

25 Primers for ATnov1

RH primers (SEQ ID NO:11) CTGCTGCCAACTAACATTGC
(SEQ ID NO:12) CACACACTAACCATGCCTCT

237 bp product

30 RT-PCR primers (SEQ ID NO:13) TCCAGTCATTGGCTTGCAC
(SEQ ID NO:14) AAGAACCAATAAGCTGCTTACT

413 bp product

Primers for ATnov2

RH primers (SEQ ID NO:15) GTGTTTGCTAGCCACCTTGA

(SEQ ID NO:16) GGCAACACTCCTCAAAGTG
196 bp product

RT-PCR primers (SEQ ID NO:17) GATGCTTCCTCTGTGCAGT
(SEQ ID NO:18) CCTTCAAGCCGAAGAAGGCT
5 259 bp product

Primers for ATnov3

RH primers (SEQ ID NO:19) AGGAGTTCTGGTCCTTCA
(SEQ ID NO:20) CAAGCTAGACTTCAGGCCTT
10 137 bp product

RT-PCR primers (SEQ ID NO:21) GAGGAATTCTAGCTCCAATATATT
(SEQ ID NO:22) GTCCTACATGACCCACGTGTG
96 bp product

15 Results

cDNA Isolation. Large-scale exon trapping was completed across a chromosome 12 cosmid library. Approximately 2400 exons were sequenced and analyzed by BLAST algorithms to identify exons with potentially interesting homologies. Two different exons, C12B_120 and C12C_151 were identified that were 87% identical to each other and ~68% identical to a cloned organic anion transporter, OATP (Kullak-Ublick et al. (1995) *Gastroenterology* 109: 1274-1282.), at the DNA level and ~78% similar at the amino acid level. Full-length cDNA clones were isolated using GeneTrapper (Gibco-BRL) from a liver cDNA library (Gibco-BRL). The resulting clones, the largest being up to 3.0 kb, were end-sequenced using vector primers. If the end sequences provided insufficient coverage of the 25 cDNA clones, a transposon approach was used to complete the sequence of the cDNA clone.

The cDNA clones isolated with C12B_120 yielded two different sequence contigs, ATnov1 and ATnov2, which were ~89% identical to each other. ATnov2 is identical to C12B_120. cDNA clones isolated with C12C_151 generated a sequence contig, ATnov3, 30 that was ~86% identical to the first two contigs. Conceptual translations yielded predicted proteins of 688-704 amino acids in length. A multiple alignment of these three proteins is shown in figure 1. These genes also show significant homology to a human organic anion

transporter OATP (~40% identity, 60% similarity) and to a human prostaglandin transporter (~32% identity, 51% similarity) over the length of the predicted proteins.

Chromosomal Localization. The exon trapped products used in the cDNA screens 5 were trapped from a chromosome 12 cosmid library, suggesting that at least ATnov2 and ATnov3 map to chromosome 12. OATP had been previously reported to map to chromosome 12 (Kullak-Ubrick et al., *supra*.) Radiation hybrid mapping was used to confirm the localization of these to chromosome 12, as well as to map them and ATnov1 to a specific region on the chromosome. The Stanford G3 panel showed linkage of all four of these 10 genes to the marker GATA91H01, which is extrapolated to a cytogenetic location of 12p12.

Expression Analysis. OATP is expressed in multiple tissues, including brain, lung, liver, kidney, and testes (Kullak-Ubrick et al., *supra*.) RT-PCR was utilized to characterize 15 the expression pattern of the novel anion transporters. This approach used RNA from 30 different tissues to generate first strand cDNA. These cDNAs were arrayed, along with control cDNA synthesis reactions without reverse transcriptase, into 96-well trays, dried and stored until needed. This resource provides for relatively high-throughput analysis of gene expression in a large set of tissues in a cost-efficient manner. RT-PCR in this fashion allows for qualitative analysis of gene expression only.

20 PCR was performed on these plates with gene-specific primers for each of the ATnov genes. ATnov1 is expressed in fetal and adult liver; ATnov2 is expressed in adult liver and mammary gland; ATnov3 is expressed in fetal liver, adult liver, brain, adipose tissue, skin, and testes.

25 The predicted positions of transmembrane domains in the ATnov3 polypeptide are as follows:

	ATnov3
Transmembrane domain 1	29-45
Transmembrane domain 2	94-104
Transmembrane domain 3	169-189
Transmembrane domain 4	207-227
Transmembrane domain 5	259-279
Transmembrane domain 6	336-356

Transmembrane domain 7	376-396
Transmembrane domain 8	410-430
Transmembrane domain 9	481-501
Transmembrane domain 10	537-557
Transmembrane domain 11	581-601
Transmembrane domain 12	627-647

These novel members of the organic anion transporter family are expressed in the liver. Based on homology to another organic anion transporter, they are likely to be present on the basolateral surface of the hepatocytes and mediate the uptake of both xenobiotics and endogenous compounds for metabolism by the cytochrome p450s, glucuronosyl transferases, and other metabolic enzymes known to be present in the liver. The ATnov genes are all expressed in the liver, with ATnov 2 and 3 also being expressed in a limited number of other tissues. The RT-PCR approach described herein has a high level of sensitivity, with the ability to detect a control transcript diluted down to an expression level equivalent to a frequency of 1/10⁷.

The map positions of these anion transporters suggest that they lie adjacent to each other on the proximal short arm of chromosome 12. The anion transporters described herein are only ~89% identical to each other at the DNA level, suggesting that these genes arose via a recombination mechanism, but have since diverged sufficiently such that it is unlikely that these genes are polymorphic within a given population.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid encoding a mammalian ATnov protein.
2. An isolated nucleic acid according to Claim 1, wherein said ATnov protein has
5 the amino acid sequence of SEQ ID NO:4, 6, 8, or 10.
3. An isolated nucleic acid according to Claim 1, wherein said ATnov protein has
an amino acid sequence that is substantially identical to the amino acid sequence of SEQ
ID NO:4, 6, 8, or 10.
- 10 4. An isolated nucleic acid according to Claim 1, comprising the nucleotide
sequence as set forth in SEQ ID NO:1,2, 3, 5, 7, or 9.
- 15 5. An isolated nucleic acid that hybridizes under stringent conditions to the
nucleic acid sequence of claim 4.
- 20 6. An expression cassette comprising a transcriptional initiation region functional
in an expression host, a nucleic acid having a sequence of the isolated nucleic acid
according to Claim 1 under the transcriptional regulation of said transcriptional initiation
region, and a transcriptional termination region functional in said expression host.
- 25 7. A cell comprising an expression cassette according to Claim 6 as part of an
extrachromosomal element or integrated into the genome of a host cell as a result of
introduction of said expression cassette into said host cell, and the cellular progeny of said
host cell.
- 30 8. A method for producing mammalian ATnov protein, said method comprising:
growing a cell according to Claim 7, whereby said mammalian ATnov protein is
expressed; and
isolating said ATnov protein free of other proteins.
9. A purified polypeptide composition comprising at least 50 weight % of the
protein present as a ATnov protein or a fragment thereof.

10. A monoclonal antibody binding specifically to an ATnov protein.
11. A non-human transgenic animal model for ATnov gene function wherein said
5 transgenic animal comprises an introduced alteration in an ATnov gene.
12. The animal model of claim 11, wherein said animal is heterozygous for said introduced alteration.
- 10 13. The animal model of claim 12, wherein said animal is homozygous for said introduced alteration.
14. The animal model of claim 12, wherein said introduced alteration is a knockout of endogenous ATnov gene expression.
- 15 15. An isolated nucleic acid probe comprising an ATnov 3' sequence polymorphism, as part of other than a naturally occurring chromosome.
16. A nucleic acid probe according to Claim 15, wherein said probe is conjugated
20 to a detectable marker.
17. An array of oligonucleotides comprising:
two or more probes for detection of ATnov3 locus polymorphisms.

SEQUENCE LISTING

<110> Miller, Andrew
 Buckler, Alan
 Laubert, Boris
 <120> Human Anion Transporter Genes
 <130> SEQ-23P
 <150> 60/095,835
 <151> 1998-08-07
 <160> 23
 <170> FastSEQ for Windows Version 3.0
 <210> 1
 <211> 2595
 <212> DNA
 <213> H. sapiens

<400> 1

aaaaaaagga	tggacttggtt	gcagttgctg	tagcattcaa	agtcaaggtg	atcatttcaa	60
accaagcata	agcaacaatt	aaaaatattc	acttggtata	tgtatgtttaa	taatggacca	120
acatcaacat	ttgaataaaa	cagcagagtc	agcatcttca	gagaaaaaaga	aaacaagacg	180
ctgcaatgg	ttcaagatgt	tcttggcagc	cctgtcattc	agctatattt	ctaaaggact	240
aggtggaaatc	attatgaaaa	tttccatcac	tcaaataagaa	aggagattt	acatatccctc	300
ttctcttgct	ggtttaattt	atggaagctt	tgaatttgg	aatttgcattt	tgattgtatt	360
tgtatgttac	tttggatcta	aactacacag	accgaagttt	attggattt	tttgcctcct	420
tatggaaact	ggaagtattt	tgacatctt	accacattt	ttcatggat	attataggta	480
ttctaaagaa	accatataat	atccatcaga	aaattcaaca	tcaagttt	caacctgttt	540
aattaatcaa	accttatcat	tcaatggaa	atcacctgag	atagtagaaa	aaagattgtgt	600
aaaggaatct	gggtcacaca	tgtggatcta	tgtcttcatt	gggaatatgc	ttcgtggcat	660
aggggaaacc	cccatagttac	cattggggat	ttcatacatt	gatgatttt	caaaagaagg	720
acattcttc	ttgtatttt	gtagttt	tgcaatagga	atgattggc	cagtcttgg	780
ctttgcactg	ggatctctgt	ttgtctaaat	gtacgtggat	attggatatg	tagatctgag	840
cactatcaga	ataactccta	aggactctcg	ttgggttgg	gcttggggc	ttggttccct	900
tgtgtctgga	ctatttcca	ttatttcttc	cataccattt	tttttcttgc	ccgaaaaatc	960
caaataaaacc	acaaaaaagaa	agaaaaattt	cactatcatt	gcatgtgt	aaaacaaaatg	1020
atgatagaaaa	tcaaacatgt	aatttgcacca	accaaggaaa	aatgttacc	aaaaatgtga	1080
ctggttttt	ccagttttt	aaaagcatcc	ttacaatcc	cctgtatgtt	atattctgc	1140
ttttgacatt	tttacaatgt	agcagttt	ttggttttt	tacttacgtc	tttaaatata	1200
tggagcaaca	gtacggtcag	tctgcatttc	atgtaactt	tttggggat	atcataacca	1260
ttcctacgg	tgcaacttgg	atgttttt	gaggatttt	cattaaaaaa	ttcaaaattgt	1320
cttttagttgg	aatttgcaaa	ttttcatttc	ttacttgcatt	gataatccctc	ttgtttcaac	1380
ttctatattt	ccctctaaatc	tgcggaaagca	aatcagtgc	cggcctaacc	ttgacctatg	1440
atggaaataa	ttcagttggca	tctcatgttag	atgttaccat	ttcttatttgc	aactcagat	1500
gcaatttgta	tgaaagtca	tggttggaccag	tctgtggaa	caatggata	acttacctgt	1560
caccttgtct	agcaggatgc	aaatcctcaa	gtggtattaa	aaagcataca	gtgtttata	1620
actgttagtt	tgtggaaatgt	actggtctcc	agaacagaaa	ttactcagca	cacttgggtg	1680
aatggccaaag	agataataact	tgtacaaggaa	aatttttcat	ctatgttgca	attcaagtca	1740
taaactctt	tttctctgca	acagggttca	ccacatttt	tttggact	gtgaagattt	1800
ttcaacactg	attgaaagca	cttgcattgg	gtttccatgt	aatgttata	agaacactaa	1860
gaggaatttc	agctccaaata	tatttttgggg	ctctgtatgt	taaaacatgt	atgaaatgggt	1920
ccaccaacaa	ctgtggagca	caaggggtt	gttaggatata	taattccgtt	ttttttggaa	1980
gggtctactt	gggttttactt	atagctttaa	gattccatgt	acttgcattt	tatattgttt	2040
tcattttgc	tatgaaagaaa	aaatttcaag	gaaaagatata	caaggcatcg	gacaatgaaa	2100
gaaaagtaat	ggatgaagca	aacttagaat	tcttaataaa	tggtaacat	tttgcacattt	2160
ctgctggaaac	agatagtaaa	acatgttaatt	tggacatgtca	agacaatgtc	gctgccaact	2220

aacattgcatt	tgattcatta	agatgttatt	tttgagggtgt	tcctggtctt	tcactgacaa	2280
ttccaacatt	ctttacttac	agtggaccaa	tggataagtc	tatgcattct	taataaaacta	2340
taaaaaatgg	gagtagccat	ggtaggata	tagctatgcc	ttttaggtt	agatttagaat	2400
atatgatcca	taaaaattt	aagtggaggg	catggttagt	gtgtgataca	ataaaaaagta	2460
atgttttgtt	agttgttaact	gctaataaaa	ccagtgcata	gaatataagg	gaggaaaaaa	2520
ggacaagata	gattaatgc	ctaaataaag	agaaaaagcct	gatgccttta	aaaaaaaaatg	2580
aaaaaaaaaa	aaaaaa					2595

<210> 2
 <211> 3273
 <212> DNA
 <213> H. sapiens

<400> 2						
cccacgcgtc	cgatcagaaa	aaggatggac	ttgttgcagt	tgctgttagca	tccaaagtca	60
agtgtatcat	ttcaaaccac	gcatcagcaa	caattaaaaa	tattcaactt	gtatctgttag	120
ttaataatg	gaccaacatc	aacatttga	taaaacagca	gagtcagcat	cttcagagaa	180
aaagaaaaca	agacgctgca	atggattca	gatgttctt	gcagccctgt	cattcagcta	240
tattgctaaa	gcaactaggt	gaatcattat	gaaaattttc	atcaactaaa	tagaaaggag	300
atttgcata	tcctcttc	ttgttggttt	aattgtgga	agctttgaaa	ttggaaattt	360
gcttgcatt	gtatgtttaa	gttactttgg	atctaaacta	cacagaccga	agttatttgg	420
aattggtt	tcctttaggg	gaactggaag	tattttaca	gctttaccac	atttcttcat	480
ggataatct	tcttgacact	cctgtcattt	agctatgtt	ctaaagcact	agctggattt	540
tttatgaaaa	tatcaaccac	tcaaatagaa	aggagattt	agatatcctc	ttcttcttgg	600
ggttaattt	atggaagctt	cgaatagga	aatttggttt	tgatttgcatt	tgtaaatttac	660
tttggatcta	aactacacag	acccaagttt	atttggattt	gttggtttct	tatgggaact	720
ggaagtattt	tgtggatctt	accacattt	ttcatggat	attacaggta	ttctaaagaa	780
accaatattt	atccatcaga	aaattcaaca	tcaaacttac	caaactgttt	attaatcaaa	840
atgttattcac	tcaatagaac	accgtcttag	ataatagaaa	gaggttgcgt	gaaggaaatct	900
gggtcacaca	tgtggatcta	tgtcttcattt	ggtatata	ttcggtggat	agggggaaacc	960
cccatagttac	cattggggat	ttcttacattt	gatgttggat	caaaaagagg	acatttcttcc	1020
ttgttatttt	gtactgtgaa	tgcataatggaa	atgactgggt	tagtttttgc	ttttatgtctg	1080
ggatctctgt	ttgtcttttt	gtatgtggat	atcgatata	tggatctgag	cactatcagg	1140
ataactctca	aggactctcg	ttgggttggaa	gcttgggtgc	ttgggttctt	tgtgtcttgg	1200
atagatccat	ttatttcttc	tataccattt	tttttcttgc	ctctaaatcc	aaataaaacca	1260
cagaaagaaaa	ggaaagttt	actatttttt	catgtgcata	aaactaatga	taaaaggaat	1320
caaatacgta	atttgcacaa	ccgaagaaaa	tatattacca	aaaatgtgac	ttggcttttc	1380
cagtcttgc	aaagcattt	taccaatccc	ctgtatgtt	tatttgcatt	ttttacatttgc	1440
ttacacatga	gcagctacat	tgtttctt	acttata	ttaaaatgtt	ggagcaacag	1500
tatgggtt	ctgcatactt	gactaacttt	ttgttggag	tcctcgccct	acctactgtt	1560
gcaattggca	tggtttcagg	aggatata	attaaaaaaat	tcaaatttgc	tttagttgg	1620
cttgccaaat	tggcatttt	ttctgcaaca	gtgcatttct	tatctcaat	tttatatttc	1680
tttctaatct	gtgaaagcaa	atcgttgc	ggcctaacc	tgacccat	tggaaatagt	1740
ccagtaatgt	ctcatgttgc	tgttaccattt	tcttatttgc	actcagatgt	caatttgcatt	1800
gaaagtcaat	ggggacccgt	ctgtggaaac	aatggaaata	cttacctgtc	accttgcata	1860
gcaggatgc	aatcttcaag	ttgttataaa	gagccat	tgtttataa	ctgtagctgt	1920
gttggaaat	ttgggttccaa	gaacaaaaat	tactcagcgc	acttgggtga	atgcccaga	1980
gatgtgtt	gtacaaggaa	atcttacgtt	tattttgcatt	ttcaatgtt	agatgttttcc	2040
ctctgtgcac	ttggacttac	ctcatatttcc	gtgttgcatt	tttagatttgc	tcaacctgaa	2100
ttgaaagcac	ttgcaatcgg	cttccatttca	atgattatgc	gatgccttgc	aggtatttctt	2160
tttccaaat	attttggggc	tctgttgcatt	acaacgttgc	tgaatgttgc	caccaacacgc	2220
tgtggagcac	gaggggttgc	taggtatata	aattccat	attttggaaag	agccttcttcc	2280
ggcttgcagg	tagccctaat	atttccatgt	cttgcattt	ttactgtatt	tatttttgtt	2340
gtaaggaaaa	aatccatgg	aaaggatacc	aaagtattt	aaaatgttgc	acaaagtaat	2400
gatgtgtt	acttgcattt	cttacacgc	agtgttgcatt	ttgttgcattt	tgctgttgc	2460
cagtaatgt	tgttataaa	aggaggaaaa	aaataat	gctgttgcattt	ccaaactat	2520
tatttgcattt	ataagacgtt	atttttgcattt	ttgttgcattt	tttgcatttgc	agaattccca	2580
cattcttac	ttatgtatgc	acaatgttgc	agcctatgaa	tttataatgc	acaaaactat	2640
aaaaaatgtt	accatgttgc	aggacatgc	tacacaagca	tttgcatttgc	agaatata	2700
attcataaaa	atttgcatttgc	agaggtatgc	ttaatgttgc	atagaagaaaa	agttacttgc	2760
tcaggtatgt	gtacttgcata	ataaaaacca	tgactgttgc	acaatgttgc	gtaaaaaggt	2820
ggagatagat	taatgcctt	aataacgcaga	gaaccttgc	cctttttaa	acaaaaccaa	2880

aaccatttagacatTTTact tagtcctaaa atctagcctg gatttatgct ataatgatat	2940
ctatTTTca tgTTaaatttacattactc agaaattata aatattatta ctTTataatt	3000
tgaaatttgc ttgcTAGCC acCTTgatgt atTTTCTTCC aaACTCCCAT taagatacta	3060
ttgaaaaaaat agaaatAGTC aaatatttgc aaggTataat tgTTAGGCAA catattatag	3120
catgtgttaa gtttCTGCTA ggCTCATGGa aattttttt ttTatTTTG ttCCATTTT	3180
attcactttt aggaagtGTT GCTTTTTT ttGATGtaCT taaatGGCTa aaataaaaaaa	3240
gacaatcaca agaaaaaaa aaaaaaaaaaaa aaa	3273

<210> 3
 <211> 2799
 <212> DNA
 <213> H. sapiens

<220>
 <221> CDS
 <222> (100)...(1729)
 <223> ATnov3.1 Coding sequence

<221> variation
 <222> (1705)...(1710)
 <223> Polymorphism of 5 or 6 thymidine residues

<221> variation
 <222> (487)...(487)
 <223> Polymorphism of A or G

<221> variation
 <222> (670)...(670)
 <223> Polymorphism of C or T

<400> 3
 gtggacttgc tgcAGTTGCT gtaggattct aaatccaggT gattgtttca aactgagcat
 caacaacaaa aacatttGta tgatatctat atttcaatc atg gac caa aat caa
 Met Asp Gln Asn Gln
 1 5

cat ttg aat aaa aca gca gag gca caa cct tca gag aat aag aaa aca
 His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser Glu Asn Lys Lys Thr
 10 15 20

aga tac tgc aat gga ttg aag atg ttc ttg gca gct ctg tca ctc agc
 Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala Ala Leu Ser Leu Ser
 25 30 35

ttt att gct aag aca cta ggt gca att att atg aaa agt tcc atc att
 Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met Lys Ser Ser Ile Ile
 40 45 50

cat ata gaa cgg aga ttt gag ata tcc tct tct ctt gtt ggt ttt att
 His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser Leu Val Gly Phe Ile
 55 60 65

gac gga agc ttt gaa att gga aat ttg ctt gtg att gta ttt gtg agt
 Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val Ile Val Phe Val Ser
 70 75 80 85

tac ttt gga tcc aaa cta cat aga cca aag tta att gga atc ggt tgc
 Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu Ile Gly Ile Gly Cys
 90 95 100

ttc att atg gga att gga ggt gtt ttg act gct ttg cca cat ttc ttc Phe Ile Met Gly Ile Gly Gly Val Ile Thr Ala Leu Pro His Phe Phe 105 110 115	450
atg gga tat tac agg tat tct aaa gaa act aat atc rat tca tca gaa Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn Ile Xaa Ser Ser Glu 120 125 130	498
aat tca aca tcg acc tta tcc act tgt tta att aat caa att tta tca Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile Asn Gln Ile Leu Ser 135 140 145	546
ctc aat aga gca tca cct gag ata gtg gga aaa ggt tgt tta aag gaa Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys Gly Cys Leu Lys Glu 150 155 160 165	594
tct ggg tca tac atg tgg ata tat gtg ttc atg ggt aat atg ctt cgt Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met Gly Asn Met Leu Arg 170 175 180	642
gga ata ggg gag act ccc ata gta cca ytg ggg ctt tct tac att gat Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly Leu Ser Tyr Ile Asp 185 190 195	690
gat ttc gct aaa gaa gga cat tct tct ttg tat tta ggt ata ttg aat Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr Leu Gly Ile Leu Asn 200 205 210	738
gca ata gca atg att ggt cca atc att ggc ttt acc ctg gga tct ctg Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe Thr Leu Gly Ser Leu 215 220 225	786
ttt tct aaa atg tac gtg gat att gga tat gta gat cta agc act atc Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val Asp Leu Ser Thr Ile 230 235 240 245	834
agg ata act cct act gat tct cga tgg gtt gga gct tgg tgg ctt aat Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly Ala Trp Trp Leu Asn 250 255 260	882
ttc ctt gtg tct gga cta ttc tcc att att tct tcc ata cca ttc ttt Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser Ser Ile Pro Phe Phe 265 270 275	930
ttc ttg ccc caa act cca aat aaa cca caa aaa gaa aga aaa gct tca Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys Glu Arg Lys Ala Ser 280 285 290	978
ctg tct ttg cat gtg ctg gaa aca aat gat gaa aag gat caa aca gct Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu Lys Asp Gln Thr Ala 295 300 305	1026
aat ttg acc aat caa gga aaa aat att acc aaa aat gtg act ggt ttt Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys Asn Val Thr Gly Phe 310 315 320 325	1074
ttc cag tct ttt aaa agc atc ctt act aat ccc ctg tat gtt atg ttt Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro Leu Tyr Val Met Phe 330 335 340	1122

gtg ctt ttg acg ttg tta caa gta agc agc tat att ggt gct ttt act	1170
Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr Ile Gly Ala Phe Thr	
345 350 355	
tat gtc ttc aaa tac gta gag caa cag tat ggt cag cct tca tct aag	1218
Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly Gln Pro Ser Ser Lys	
360 365 370	
gct aac atc tta ttg gga gtc ata acc ata cct att ttt gca agt gga	1266
Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro Ile Phe Ala Ser Gly	
375 380 385	
atg ttt tta gga gga tat atc att aaa aaa ttc aaa ctg aac acc gtt	1314
Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe Lys Leu Asn Thr Val	
390 395 400 405	
gga att gcc aaa ttc tca tgt ttt act gct gtg atg tca ttg tcc ttt	1362
Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val Met Ser Leu Ser Phe	
410 415 420	
tac cta tta tat ttt ttc ata ctc tgt gaa aac aaa tca gtt gcc gga	1410
Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn Lys Ser Val Ala Gly	
425 430 435	
cta acc atg acc tat gat gga aat aat cca gtg aca tct cat aga gat	1458
Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val Thr Ser His Arg Asp	
440 445 450	
gta cca ctt tct tat tgc aac tca gac tgc aat tgt gat gaa agt caa	1506
Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn Cys Asp Glu Ser Gln	
455 460 465	
tgg gaa cca gtc tgt gga aac aat gga ata act tac atc tca ccc tgt	1554
Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr Tyr Ile Ser Pro Cys	
470 475 480 485	
cta gca ggt tgc aaa tct tca agt ggc aat aaa aag cct ata gtg ttt	1602
Leu Ala Gly Cys Lys Ser Ser Gly Asn Lys Lys Pro Ile Val Phe	
490 495 500	
tac aac tgc agt tgt ttg gaa gta act ggt ctc cag aac aga aat tac	1650
Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu Gln Asn Arg Asn Tyr	
505 510 515	
tca gcc cat ttg ggt gaa tgc cca aga gat gat gct tgt aca agg aaa	1698
Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp Ala Cys Thr Arg Lys	
520 525 530	
ttt tac ttt ttg ttg caa tac aag tct tga a ttatatttc tctgcacttg	1749
Phe Tyr Phe Leu Leu Gln Tyr Lys Ser *	
535 540	
gaggcacctc acatgtcatg ctgattgtta aaattgttca acctgaattg aaatcacttg	1809
cactgggtt ccaactcaatg gtatacggag cacttagggg aattcttagct ccaatataatt	1869
ttggggctct gattgataca acgtgtataa agtggtccac caacaactgt ggcacacgtg	1929
ggtcatgttag gacatataat tccacatcat ttcaagggt ctacttggc ttgtcttcaa	1989
tgttaagagt ctcatcaat ttatattaaat ttatgccatg aaaaaaaaaat	2049
atcaagagaa agatatacat gcatcagaaa atggaagtgt catggatgaa gcaaaacttag	2109
aatccttaaa taaaaataaa cattttgtcc cttctgtgg ggcagatgt gaaacacatt	2169
gttaagggga gaaaaaaaaagc cacttctgct tctgtgttca caaacagcat tgcattgatt	2229
cagtaagatg ttattttga ggagttctg gtccttcac taagaattc cacatcttt	2289
atggtggaaag tataaataag cctatgaact tataataaaa caaactgttag gtagaaaaaa	2349

tgagagtact cattgttaca ttatagctac atatttgtgg ttaaggtag actatatgtat	2409
ccatacaaat taaaagtgaga gacatggta ctgtgtataa aaagaaaaaa tacttggta	2469
ggtaattctta attcttaataa aaacaaatgaa gtatcataca ggttagaggtt aaaaaggagg	2529
agcttagatccatataa taaagagaaa tgccttagtgtt ctattttttt aaacaaacaa	2589
acacagagtt tgaactataa tactaaggcc tgaagtcttag cttggatata tgctacaata	2649
atatctgtta ctcacataaa attatattt tcacagactt tatcaatgtta taattaacaa	2709
ttatcttggta taagtaaatt tagaatacat ttaagtattt tgaaagaaat aaagacattc	2769
caatatttgc aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa	2799

<210> 4
 <211> 542
 <212> PRT
 <213> H. sapiens

<220>
 <221> VARIANT
 <222> (130)...(130)
 <223> Xaa = Asp or Asn

<221> VARIANT
 <222> (191)...(191)
 <223> Xaa = Leu

<400> 4

Met Asp Gln Asn Gln His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser	
1 5 10 15	
Glu Asn Lys Lys Thr Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala	
20 25 30	
Ala Leu Ser Leu Ser Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met	
35 40 45	
Lys Ser Ser Ile Ile His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser	
50 55 60	
Leu Val Gly Phe Ile Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val	
65 70 75 80	
Ile Val Phe Val Ser Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu	
85 90 95	
Ile Gly Ile Gly Cys Phe Ile Met Gly Ile Gly Gly Val Leu Thr Ala	
100 105 110	
Leu Pro His Phe Phe Met Gly Tyr Arg Tyr Ser Lys Glu Thr Asn	
115 120 125	
Ile Xaa Ser Ser Glu Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile	
130 135 140	
Asn Gln Ile Leu Ser Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys	
145 150 155 160	
Gly Cys Leu Lys Glu Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met	
165 170 175	
Gly Asn Met Leu Arg Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly	
180 185 190	
Leu Ser Tyr Ile Asp Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr	
195 200 205	
Leu Gly Ile Leu Asn Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe	
210 215 220	
Thr Leu Gly Ser Leu Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val	
225 230 235 240	
Asp Leu Ser Thr Ile Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly	
245 250 255	
Ala Trp Trp Leu Asn Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser	
260 265 270	
Ser Ile Pro Phe Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys	
275 280 285	
Glu Arg Lys Ala Ser Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu	
290 295 300	

Lys Asp Gln Thr Ala Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys
 305 310 315 320
 Asn Val Thr Gly Phe Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro
 325 330 335
 Leu Tyr Val Met Phe Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr
 340 345 350
 Ile Gly Ala Phe Thr Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly
 355 360 365
 Gln Pro Ser Ser Lys Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro
 370 375 380
 Ile Phe Ala Ser Gly Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe
 385 390 395 400
 Lys Leu Asn Thr Val Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val
 405 410 415
 Met Ser Leu Ser Phe Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn
 420 425 430
 Lys Ser Val Ala Gly Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val
 435 440 445
 Thr Ser His Arg Asp Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn
 450 455 460
 Cys Asp Glu Ser Gln Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr
 465 470 475 480
 Tyr Ile Ser Pro Cys Leu Ala Gly Cys Lys Ser Ser Ser Gly Asn Lys
 485 490 495
 Lys Pro Ile Val Phe Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu
 500 505 510
 Gln Asn Arg Asn Tyr Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp
 515 520 525
 Ala Cys Thr Arg Lys Phe Tyr Phe Leu Leu Gln Tyr Lys Ser
 530 535 540

<210> 5
 <211> 2800
 <212> DNA
 <213> H. sapiens

<220>
 <221> CDS
 <222> (100)...(2175)
 <223> Coding sequence of ATnov3.1

<221> variation
 <222> (1705)...(1710)
 <223> Polymorphism of 5 or 6 T residues.

<221> variation
 <222> (487)...(487)
 <223> Polymorphism of A or G

<221> variation
 <222> (670)...(670)
 <223> Polymorphism of C or T

<400> 5
 gtggacttgt tgcatgttgcgtt gtaggattctt aaatccagggtt gattgtttca aactgagcat 60
 caacaacaaa aacatttgcgtt tgatatctat atttcaatc atg gac caa aat caa 114
 Met Asp Gln Asn Gln
 1 5

cat ttg aat aaa aca gca gag gca caa cct tca gag aat aag aaa aca 162
 His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser Glu Asn Lys Lys Thr
 10 15 20

aga tac tgc aat gga ttg aag atg ttc ttg gca gct ctg tca ctc agc Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala Ala Leu Ser Leu Ser 25 30 35	210
ttt att gct aag aca cta ggt gca att att atg aaa agt tcc atc att Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met Lys Ser Ser Ile Ile 40 45 50	258
cat ata gaa cgg aga ttt gag ata tcc tct tct ctt gtt ggt ttt att His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser Leu Val Gly Phe Ile 55 60 65	306
gac gga agc ttt gaa att gga aat ttg ctt gtg att gta ttt gtg agt Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val Ile Val Phe Val Ser 70 75 80 85	354
tac ttt gga tcc aaa cta cat aga cca aag tta att gga atc ggt tgt Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu Ile Gly Ile Gly Cys 90 95 100	402
ttc att atg gga att gga ggt gtt ttg act gct ttg cca cat ttc ttc Phe Ile Met Gly Ile Gly Val Leu Thr Ala Leu Pro His Phe Phe 105 110 115	450
atg gga tat tac agg tat tct aaa gaa act aat atc rat tca tca gaa Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn Ile Xaa Ser Ser Glu 120 125 130	498
aat tca aca tcg acc tta tcc act tgt tta att aat caa att tta tca Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile Asn Gln Ile Leu Ser 135 140 145	546
ctc aat aga gca tca cct gag ata gtg gga aaa ggt tgt tta aag gaa Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys Gly Cys Leu Lys Glu 150 155 160 165	594
tct ggg tca tac atg tgg ata tat gtg ttc atg ggt aat atg ctt cgt Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met Gly Asn Met Leu Arg 170 175 180	642
gga ata ggg gag act ccc ata gta cca ytg ggg ctt tct tac att gat Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly Leu Ser Tyr Ile Asp 185 190 195	690
gat ttc gct aaa gaa gga cat tct tct ttg tat tta ggt ata ttg aat Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr Leu Gly Ile Leu Asn 200 205 210	738
gca ata gca atg att ggt cca atc att ggc ttt acc ctg gga tct ctg Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe Thr Leu Gly Ser Leu 215 220 225	786
ttt tct aaa atg tac gtg gat att gga tat gta gat cta agc act atc Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val Asp Leu Ser Thr Ile 230 235 240 245	834
agg ata act cct act gat tct cga tgg gtt gga gct tgg tgg ctt aat Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly Ala Trp Trp Leu Asn 250 255 260	882

ttc ctt gtg tct gga cta ttc tcc att att tct tcc ata cca ttc ttt Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser Ser Ile Pro Phe Phe 265 270 275	930
ttc ttg ccc caa act cca aat aaa cca caa aaa gaa aga aaa gct tca Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys Glu Arg Lys Ala Ser 280 285 290	978
ctg tct ttg cat gtg ctg gaa aca aat gat gaa aag gat caa aca gct Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu Lys Asp Gln Thr Ala 295 300 305	1026
aat ttg acc aat caa gga aaa aat att acc aaa aat gtg act ggt ttt Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys Asn Val Thr Gly Phe 310 315 320 325	1074
ttc cag tct ttt aaa agc atc ctt act aat ccc ctg tat gtt atg ttt Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro Leu Tyr Val Met Phe 330 335 340	1122
gtg ctt ttg acg ttg tta caa gta agc agc tat att ggt gct ttt act Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr Ile Gly Ala Phe Thr 345 350 355	1170
tat gtc ttc aaa tac gta gag caa cag tat ggt cag cct tca tct aag Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly Gln Pro Ser Ser Lys 360 365 370	1218
gct aac atc tta ttg gga gtc ata acc ata cct att ttt gca agt gga Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro Ile Phe Ala Ser Gly 375 380 385	1266
atg ttt tta gga gga tat atc att aaa aaa ttc aaa ctg aac acc gtt Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe Lys Leu Asn Thr Val 390 395 400 405	1314
gga att gcc aaa ttc tca tgt ttt act gct gtg atg tca ttg tcc ttt Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val Met Ser Leu Ser Phe 410 415 420	1362
tac cta tta tat ttt ttc ata ctc tgt gaa aac aaa tca gtt gcc gga Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn Lys Ser Val Ala Gly 425 430 435	1410
cta acc atg acc tat gat gga aat aat cca gtg aca tct cat aga gat Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val Thr Ser His Arg Asp 440 445 450	1458
gta cca ctt tct tat tgc aac tca gac tgc aat tgt gat gaa agt caa Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn Cys Asp Glu Ser Gln 455 460 465	1506
tgg gaa cca gtc tgt gga aac aat gga ata act tac atc tca ccc tgt Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr Tyr Ile Ser Pro Cys 470 475 480 485	1554
cta gca ggt tgc aaa tct tca agt ggc aat aaa aag cct ata gtg ttt Leu Ala Gly Cys Lys Ser Ser Ser Gly Asn Lys, Lys Pro Ile Val Phe 490 495 500	1602

tac aac tgc agt tgt ttg gaa gta act ggt ctc cag aac aga aat tac Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu Gln Asn Arg Asn Tyr 505 510 515	1650
tca gcc cat ttg ggt gaa tgc cca aga gat gat gct tgt aca agg aaa Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp Ala Cys Thr Arg Lys 520 525 530	1698
ttt tac ttt ttg gca ata caa gtc ttg aat tta ttt ttc tct gca Phe Tyr Phe Val Ala Ile Gln Val Leu Asn Leu Phe Phe Ser Ala 535 540 545	1746
ctt gga ggc acc tca cat gtc atg ctg att gtt aaa att gtt caa cct Leu Gly Gly Thr Ser His Val Met Leu Ile Val Lys Ile Val Gln Pro 550 555 560 565	1794
gaa ttg aaa tca ctt gca ctg ggt ttc cac tca atg gtt ata cga gca Glu Leu Lys Ser Leu Ala Leu Gly Phe His Ser Met Val Ile Arg Ala 570 575 580	1842
cta gga gga att cta gct cca ata tat ttt ggg gct ctg att gat aca Leu Gly Gly Ile Leu Ala Pro Ile Tyr Phe Gly Ala Leu Ile Asp Thr 585 590 595	1890
acg tgt ata aag tgg tcc acc aac aac tgt ggc aca cgt ggg tca tgt Thr Cys Ile Lys Trp Ser Thr Asn Asn Cys Gly Thr Arg Gly Ser Cys 600 605 610	1938
agg aca tat aat tcc aca tca ttt tca agg gtc tac ttg ggc ttg tct Arg Thr Tyr Asn Ser Thr Ser Phe Ser Arg Val Tyr Leu Gly Leu Ser 615 620 625	1986
tca atg tta aga gtc tca tca ctt gtt tta tat att ata tta att tat Ser Met Leu Arg Val Ser Ser Leu Val Leu Tyr Ile Ile Leu Ile Tyr 630 635 640 645	2034
gcc atg aag aaa aaa tat caa gag aaa gat atc aat gca tca gaa aat Ala Met Lys Lys Tyr Gln Glu Lys Asp Ile Asn Ala Ser Glu Asn 650 655 660	2082
gga agt gtc atg gat gaa gca aac tta gaa tcc tta aat aaa aat aaa Gly Ser Val Met Asp Glu Ala Asn Leu Glu Ser Leu Asn Lys Asn Lys 665 670 675	2130
cat ttt gtc cct tct gct ggg gca gat agt gaa aca cat tgt taa His Phe Val Pro Ser Ala Gly Ala Asp Ser Glu Thr His Cys *680 685 690	2175
ggggagaaaa aaagccactt ctgcttctgt gtttccaaac agcattgcat tgattcagta agatgttatt tttgaggagt tcctggctt ttcactaaga atttccacat cttttatgg ggaagtataa ataagctat gaacttataa taaaacaac tgttaggtaga aaaaatgaga gtactcattt gttacattt gttacatatt tttgggttaag gtttagactat atgtatccata caaattaaag tgagagacat gttactgtg taataaaaga aaaaataactt gttcaggtaa ttcttaattct taataaaaca aatgagtatc atacaggtag aggtaaaaaa ggaggagcta gattcatata ctaagtaaag agaaatgcct agtgtctatt ttataaaaca aacaaacaca gagtttgaac tataatacta aggcctgaag tctagctgg tataatgcta caataatatc tgttactcac ataaaattat atatttcaca gactttatca atgtataatt aacaattatc ttgtttaagt aaattttagaa tacattnaag tattgtggaa gaaataaaaga cattccaata tttgcaaaaa aaaaaaaaaa aaaaa	2235 2295 2355 2415 2475 2535 2595 2655 2715 2775 2800

<210> 6
<211> 691

<212> PRT
 <213> H. sapiens

<220>
 <221> VARIANT
 <222> (130)...(130)
 <223> Xaa = Asp or Asn

<221> VARIANT
 <222> (191)...(191)
 <223> Xaa = Leu

<400> 6

Met Asp Gln Asn Gln His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser
 1 5 10 15
 Glu Asn Lys Lys Thr Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala
 20 25 30
 Ala Leu Ser Leu Ser Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met
 35 40 45
 Lys Ser Ser Ile Ile His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser
 50 55 60
 Leu Val Gly Phe Ile Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val
 65 70 75 80
 Ile Val Phe Val Ser Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu
 85 90 95
 Ile Gly Ile Gly Cys Phe Ile Met Gly Ile Gly Gly Val Leu Thr Ala
 100 105 110
 Leu Pro His Phe Phe Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn
 115 120 125
 Ile Xaa Ser Ser Glu Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile
 130 135 140
 Asn Gln Ile Leu Ser Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys
 145 150 155 160
 Gly Cys Leu Lys Glu Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met
 165 170 175
 Gly Asn Met Leu Arg Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly
 180 185 190
 Leu Ser Tyr Ile Asp Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr
 195 200 205
 Leu Gly Ile Leu Asn Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe
 210 215 220
 Thr Leu Gly Ser Leu Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val
 225 230 235 240
 Asp Leu Ser Thr Ile Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly
 245 250 255
 Ala Trp Trp Leu Asn Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser
 260 265 270
 Ser Ile Pro Phe Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys
 275 280 285
 Glu Arg Lys Ala Ser Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu
 290 295 300
 Lys Asp Gln Thr Ala Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys
 305 310 315 320
 Asn Val Thr Gly Phe Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro
 325 330 335
 Leu Tyr Val Met Phe Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr
 340 345 350
 Ile Gly Ala Phe Thr Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly
 355 360 365
 Gln Pro Ser Ser Lys Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro
 370 375 380

Ile Phe Ala Ser Gly Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe
 385 390 395 400
 Lys Leu Asn Thr Val Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val
 405 410 415
 Met Ser Leu Ser Phe Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn
 420 425 430
 Lys Ser Val Ala Gly Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val
 435 440 445
 Thr Ser His Arg Asp Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn
 450 455 460
 Cys Asp Glu Ser Gln Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr
 465 470 475 480
 Tyr Ile Ser Pro Cys Leu Ala Gly Cys Lys Ser Ser Ser Gly Asn Lys
 485 490 495
 Lys Pro Ile Val Phe Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu
 500 505 510
 Gln Asn Arg Asn Tyr Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp
 515 520 525
 Ala Cys Thr Arg Lys Phe Tyr Phe Val Ala Ile Gln Val Leu Asn
 530 535 540
 Leu Phe Phe Ser Ala Leu Gly Gly Thr Ser His Val Met Leu Ile Val
 545 550 555 560
 Lys Ile Val Gln Pro Glu Leu Lys Ser Leu Ala Leu Gly Phe His Ser
 565 570 575
 Met Val Ile Arg Ala Leu Gly Gly Ile Leu Ala Pro Ile Tyr Phe Gly
 580 585 590
 Ala Leu Ile Asp Thr Thr Cys Ile Lys Trp Ser Thr Asn Asn Cys Gly
 595 600 605
 Thr Arg Gly Ser Cys Arg Thr Tyr Asn Ser Thr Ser Phe Ser Arg Val
 610 615 620
 Tyr Leu Gly Leu Ser Ser Met Leu Arg Val Ser Ser Leu Val Leu Tyr
 625 630 635 640
 Ile Ile Leu Ile Tyr Ala Met Lys Lys Tyr Gln Glu Lys Asp Ile
 645 650 655
 Asn Ala Ser Glu Asn Gly Ser Val Met Asp Glu Ala Asn Leu Glu Ser
 660 665 670
 Leu Asn Lys Asn Lys His Phe Val Pro Ser Ala Gly Ala Asp Ser Glu
 675 680 685
 Thr His Cys
 690

<210> 7
 <211> 2360
 <212> DNA
 <213> H. sapiens

<220>
 <221> CDS
 <222> (100)...(1729)
 <223> Coding sequence ATnov3.2

<221> variation
 <222> (1705)...(1710)
 <223> Polymorphism of 5 or 6 T residues

<221> variation
 <222> (487)...(487)
 <223> Polymorphism of A or G

<221> variation
 <222> (670)...(670)
 <223> Polymorphism of C or T

<400> 7

gtggacttgt	tgca	ttgtct	aaatccaggt	gattttca	aactgagcat	60										
caacaacaaa	aacatttga	tgatatctat	atttcaatc	atg	gac	caa	aat	caa	114							
				Met	Asp	Gln	Asn	Gln								
				1				5								
cat	ttg	aat	aaa	aca	gca	gag	gca	caa	cct	tca	gag	aat	aag	aaa	aca	162
His	Leu	Asn	Lys	Thr	Ala	Glu	Ala	Gln	Pro	Ser	Glu	Asn	Lys	Lys	Thr	
				10				15					20			
aga	tac	tgc	aat	gga	ttg	aag	atg	ttc	ttg	gca	gct	ctg	tca	ctc	agc	210
Arg	Tyr	Cys	Asn	Gly	Leu	Lys	Met	Phe	Leu	Ala	Ala	Leu	Ser	Leu	Ser	
				25				30					35			
ttt	att	gct	aag	aca	cta	ggt	gca	att	att	atg	aaa	agt	tcc	atc	att	258
Phe	Ile	Ala	Lys	Thr	Leu	Gly	Ala	Ile	Ile	Met	Lys	Ser	Ser	Ile	Ile	
				40				45					50			
cat	ata	gaa	cg	aga	ttt	gag	ata	tcc	tct	tct	ctt	gtt	ggt	ttt	att	306
His	Ile	Glu	Arg	Arg	Phe	Glu	Ile	Ser	Ser	Ser	Leu	Val	Gly	Phe	Ile	
				55				60				65				
gac	gga	agc	ttt	gaa	att	gga	aat	ttg	ctt	gtg	att	gta	ttt	gtg	agt	354
Asp	Gly	Ser	Phe	Glu	Ile	Gly	Asn	Leu	Leu	Val	Ile	Val	Phe	Val	Ser	
				70				75				80		85		
tac	ttt	gga	tcc	aaa	cta	cat	aga	cca	aag	tta	att	gga	atc	ggt	tgt	402
Tyr	Phe	Gly	Ser	Lys	Leu	His	Arg	Pro	Lys	Leu	Ile	Gly	Ile	Gly	Cys	
				90				95				100				
ttc	att	atg	gga	att	gga	ggt	gtt	ttg	act	gct	ttg	cca	cat	ttc	ttc	450
Phe	Ile	Met	Gly	Ile	Gly	Gly	Val	Leu	Thr	Ala	Leu	Pro	His	Phe	Phe	
				105				110				115				
atg	gga	tat	tac	agg	tat	tct	aaa	gaa	act	aat	atc	rat	tca	tca	gaa	498
Met	Gly	Tyr	Tyr	Arg	Tyr	Ser	Lys	Glu	Thr	Asn	Ile	Xaa	Ser	Ser	Glu	
				120				125				130				
aat	tca	aca	tcg	acc	tta	tcc	act	tgt	tta	att	aat	caa	att	tta	tca	546
Asn	Ser	Thr	Ser	Thr	Leu	Ser	Thr	Cys	Leu	Ile	Asn	Gln	Ile	Leu	Ser	
				135				140				145				
ctc	aat	aga	gca	tca	cct	gag	ata	gtg	gga	aaa	ggt	tgt	tta	aag	gaa	594
Leu	Asn	Arg	Ala	Ser	Pro	Glu	Ile	Val	Gly	Lys	Gly	Cys	Leu	Lys	Glu	
				150				155				160		165		
tct	ggg	tca	tac	atg	tgg	ata	tat	gtg	ttc	atg	ggt	aat	atg	ctt	cgt	642
Ser	Gly	Ser	Tyr	Met	Trp	Ile	Tyr	Val	Phe	Met	Gly	Asn	Met	Leu	Arg	
				170				175				180				
gga	ata	ggg	gag	act	ccc	ata	gta	cca	ytg	ggg	ctt	tct	tac	att	gat	690
Gly	Ile	Gly	Glu	Thr	Pro	Ile	Val	Pro	Xaa	Gly	Leu	Ser	Tyr	Ile	Asp	
				185				190				195				
gat	ttc	gct	aaa	gaa	gga	cat	tct	tct	ttg	tat	tta	ggt	ata	ttg	aat	738
Asp	Phe	Ala	Lys	Glu	Gly	His	Ser	Ser	Leu	Tyr	Leu	Gly	Ile	Leu	Asn	
				200				205				210				
gca	ata	gca	atg	att	ggt	cca	atc	att	ggc	ttt	acc	ctg	gga	tct	ctg	786
Ala	Ile	Ala	Met	Ile	Gly	Pro	Ile	Ile	Gly	Phe	Thr	Leu	Gly	Ser	Leu	
				215				220				225				

ttt tct aaa atg tac gtg gat att gga tat gta gat cta agc act atc Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val Asp Leu Ser Thr Ile 230 235 240 245	834
agg ata act cct act gat tct cga tgg gtt gga gct tgg tgg ctt aat Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly Ala Trp Trp Leu Asn 250 255 260	882
ttc ctt gtg tct gga cta ttc tcc att att tct tcc ata cca ttc ttt Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser Ser Ile Pro Phe Phe 265 270 275	930
ttc ttg ccc caa act cca aat aaa cca caa aaa gaa aga aaa gct tca Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys Glu Arg Lys Ala Ser 280 285 290	978
ctg tct ttg cat gtg ctg gaa aca aat gat gaa aag gat caa aca gct Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu Lys Asp Gln Thr Ala 295 300 305	1026
aat ttg acc aat caa gga aaa aat att acc aaa aat gtg act ggt ttt Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys Asn Val Thr Gly Phe 310 315 320 325	1074
ttc cag tct ttt aaa agc atc ctt act aat ccc ctg tat gtt atg ttt Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro Leu Tyr Val Met Phe 330 335 340	1122
gtg ctt ttg acg ttg tta caa gta agc agc tat att ggt gct ttt act Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr Ile Gly Ala Phe Thr 345 350 355	1170
tat gtc ttc aaa tac gta gag caa cag tat ggt cag cct tca tct aag Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly Gln Pro Ser Ser Lys 360 365 370	1218
gct aac atc tta ttg gga gtc ata acc ata cct att ttt gca agt gga Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro Ile Phe Ala Ser Gly 375 380 385	1266
atg ttt tta gga gga tat atc att aaa aaa ttc aaa ctg aac acc gtt Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe Lys Leu Asn Thr Val 390 395 400 405	1314
gga att gcc aaa ttc tca tgt ttt act gct gtg atg tca ttg tcc ttt Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val Met Ser Leu Ser Phe 410 415 420	1362
tac cta tta tat ttt ttc ata ctc tgt gaa aac aaa tca gtt gcc gga Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn Lys Ser Val Ala Gly 425 430 435	1410
cta acc atg acc tat gat gga aat aat cca gtg aca tct cat aga gat Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val Thr Ser His Arg Asp 440 445 450	1458
gta cca ctt tct tat tgc aac tca gac tgc aat tgt gat gaa agt caa Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn Cys Asp Glu Ser Gln 455 460 465	1506

tgg gaa cca gtc tgt gga aac aat gga ata act tac atc tca ccc tgt	1554
Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr Tyr Ile Ser Pro Cys	
470 475 480 485	
ct a gca ggt tgc aaa tct tca agt ggc aat aaa aag cct ata gtg ttt	1602
Leu Ala Gly Cys Lys Ser Ser Gly Asn Lys Lys Pro Ile Val Phe	
490 495 500	
tac aac tgc agt tgt ttg gaa gta act ggt ctc cag aac aga aat tac	1650
Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu Gln Asn Arg Asn Tyr	
505 510 515	
tca gcc cat ttg ggt gaa tgc cca aga gat gat gct tgt aca agg aaa	1698
Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp Ala Cys Thr Arg Lys	
520 525 530	
ttt tac ttt ttg ttg caa tac aag tct tga a tttattttc tctgcacttg	1749
Phe Tyr Phe Leu Leu Gln Tyr Lys Ser *	
535 540	
gaggcacctc acatgtcatg ctgattgtta aaattgttca acctgaattg aaatcacttg	1809
cactgggtt ccactcaatg gttatacgag cactaggagg aattctagct ccaatatatt	1869
ttggggctct gattgataca acgtgtataa agtggtccac caacaactgt ggcacacgtg	1929
ggtcatgtag gacatataat tccacatcat tttcaagggg ctacttgggc ttgtcttcaa	1989
tgttaagagt ctcatcaatt gtttatata ttatattaaat ttatgccatg aagaaaaaaat	2049
atcaagagaa agatatacat gcatcagaaa atggaagtgt catggatgaa gcaaacttag	2109
aatccttaaa taaaataaa cattttgtcc cttctgctgg ggcagatagt gaaacacatt	2169
gttaagggga gaaaaaaagc cacttctgct tctgttttc caaacagcat tgcattgatt	2229
cagtaagatg ttattttga ggagttcctg gtcctttcac taagaatttc cacatcttt	2289
atggtggaaag tataaataag cctatgaact tataataaaa caaactgttag gtagaaaaaa	2349
aaaaaaaaa a	2360
<210> 8	
<211> 542	
<212> PRT	
<213> H. sapiens	
<220>	
<221> VARIANT	
<222> (130)...(130)	
<223> Xaa = Asp or Asn	
<221> VARIANT	
<222> (191)...(191)	
<223> Xaa' = Leu	
<400> 8	
Met Asp Gln Asn Gln His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser	
1 5 10 15	
Glu Asn Lys Lys Thr Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala	
20 25 30	
Ala Leu Ser Leu Ser Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met	
35 40 45	
Lys Ser Ser Ile Ile His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser	
50 55 60	
Leu Val Gly Phe Ile Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val	
65 70 75 80	
Ile Val Phe Val Ser Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu	
85 90 95	
Ile Gly Ile Gly Cys Phe Ile Met Gly Ile Gly Gly Val Leu Thr Ala	
100 105 110	

Leu Pro His Phe Phe Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn
 115 120 125
 Ile Xaa Ser Ser Glu Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile
 130 135 140

Asn Gln Ile Leu Ser Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys
 145 150 155 160
 Gly Cys Leu Lys Glu Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met
 165 170 175
 Gly Asn Met Leu Arg Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly
 180 185 190
 Leu Ser Tyr Ile Asp Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr
 195 200 205
 Leu Gly Ile Leu Asn Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe
 210 215 220
 Thr Leu Gly Ser Leu Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val
 225 230 235 240
 Asp Leu Ser Thr Ile Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly
 245 250 255
 Ala Trp Trp Leu Asn Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser
 260 265 270
 Ser Ile Pro Phe Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys
 275 280 285
 Glu Arg Lys Ala Ser Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu
 290 295 300
 Lys Asp Gln Thr Ala Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys
 305 310 315 320
 Asn Val Thr Gly Phe Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro
 325 330 335
 Leu Tyr Val Met Phe Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr
 340 345 350
 Ile Gly Ala Phe Thr Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly
 355 360 365
 Gln Pro Ser Ser Lys Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro
 370 375 380
 Ile Phe Ala Ser Gly Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe
 385 390 395 400
 Lys Leu Asn Thr Val Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val
 405 410 415
 Met Ser Leu Ser Phe Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn
 420 425 430
 Lys Ser Val Ala Gly Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val
 435 440 445
 Thr Ser His Arg Asp Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn
 450 455 460
 Cys Asp Glu Ser Gln Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr
 465 470 475 480
 Tyr Ile Ser Pro Cys Leu Ala Gly Cys Lys Ser Ser Ser Gly Asn Lys
 485 490 495
 Lys Pro Ile Val Phe Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu
 500 505 510
 Gln Asn Arg Asn Tyr Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp
 515 520 525
 Ala Cys Thr Arg Lys Phe Tyr Phe Leu Leu Gln Tyr Lys Ser
 530 535 540

<210> 9
 <211> 2361
 <212> DNA
 <213> H. sapiens

<220>

<221> CDS
 <222> (100)...(2175)
 <223> Coding sequence ATnov3.2

<221> variation
 <222> (1705)...(1710)
 <223> Polymorphism of 5 or 6 T residues

<221> variation
 <222> (487)...(487)
 <223> Polymorphism of A or G residue

<221> variation
 <222> (670)...(670)
 <223> Polymorphism of C or T residue.

<400> 9

gtggacttgt	tgcagttgct	gtaggattct	aaatccaggt	gattgtttca	aactgagcat	60
caacaacaaa	aacatttcta	tgatatctat	atttcaatc	atg gac caa aat caa		114
				Met Asp Gln Asn Gln		
				1	5	
cat ttg aat aaa aca gca gag gca caa cct tca gag aat aag aaa aca						162
His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser Glu Asn Lys Lys Thr						
10		15		20		
aga tac tgc aat gga ttg aag atg ttc ttg gca gct ctg tca ctc agc						210
Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala Ala Leu Ser Leu Ser						
25		30		35		
ttt att gct aag aca cta ggt gca att att atg aaa agt tcc atc att						258
Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met Lys Ser Ser Ile Ile						
40		45		50		
cat ata gaa cgg aga ttt gag ata tcc tct tct gtt ggt ttt att						306
His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser Leu Val Gly Phe Ile						
55		60		65		
gac gga agc ttt gaa att gga aat ttg ctt gtg att gta ttt gtg agt						354
Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val Ile Val Phe Val Ser						
70		75		80		85
tac ttt gga tcc aaa cta cat aga cca aag tta att gga atc ggt tgt						402
Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu Ile Gly Ile Gly Cys						
90		95		100		
ttc att atg gga att gga ggt gtt ttg act gct ttg cca cat ttc ttc						450
Phe Ile Met Gly Ile Gly Val Leu Thr Ala Leu Pro His Phe Phe						
105		110		115		
atg gga tat tac agg tat tct aaa gaa act aat atc rat tca tca gaa						498
Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn Ile Xaa Ser Ser Glu						
120		125		130		
aat tca aca tcg acc tta tcc act tgt tta att aat caa att tta tca						546
Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile Asn Gln Ile Leu Ser						
135		140		145		
ctc aat aga gca tca cct gag ata gtg gga aaa ggt tgt tta aag gaa						594
Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys Gly Cys Leu Lys Glu						
150		155		160		165

tct	ggg	tca	tac	atg	tgg	ata	tat	gtg	ttc	atg	ggt	aat	atg	ctt	cgt	642
Ser	Gly	Ser	Tyr	Met	Trp	Ile	Tyr	Val	Phe	Met	Gly	Asn	Met	Leu	Arg	
170									175				180			
gga	ata	ggg	gag	act	ccc	ata	gta	cca	ytg	ggg	ctt	tct	tac	att	gat	690
Gly	Ile	Gly	Glu	Thr	Pro	Ile	Val	Pro	Xaa	Gly	Leu	Ser	Tyr	Ile	Asp	
185									190				195			
gat	ttc	gct	aaa	gaa	gga	cat	tct	tct	ttg	tat	tta	ggt	ata	ttg	aat	738
Asp	Phe	Ala	Lys	Glu	Gly	His	Ser	Ser	Leu	Tyr	Leu	Gly	Ile	Leu	Asn	
200							205					210				
gca	ata	gca	atg	att	ggt	cca	atc	att	ggc	ttt	acc	ctg	gga	tct	ctg	786
Ala	Ile	Ala	Met	Ile	Gly	Pro	Ile	Ile	Gly	Phe	Thr	Leu	Gly	Ser	Leu	
215							220					225				
ttt	tct	aaa	atg	tac	gtg	gat	att	gga	tat	gta	gat	cta	agc	act	atc	834
Phe	Ser	Lys	Met	Tyr	Val	Asp	Ile	Gly	Tyr	Val	Asp	Leu	Ser	Thr	Ile	
230						235				240		245				
agg	ata	act	cct	act	gat	tct	cga	tgg	gtt	gga	gct	tgg	tgg	ctt	aat	882
Arg	Ile	Thr	Pro	Thr	Asp	Ser	Arg	Trp	Val	Gly	Ala	Trp	Trp	Leu	Asn	
250						255				260						
ttc	ctt	gtg	tct	gga	cta	ttc	tcc	att	att	tct	tcc	ata	cca	ttc	ttt	930
Phe	Leu	Val	Ser	Gly	Leu	Phe	Ser	Ile	Ile	Ser	Ser	Ile	Pro	Phe	Phe	
265						270				275						
ttc	ttg	ccc	caa	act	cca	aat	aaa	cca	caa	aaa	gaa	aga	aaa	gct	tca	978
Phe	Leu	Pro	Gln	Thr	Pro	Asn	Lys	Pro	Gln	Lys	Glu	Arg	Lys	Ala	Ser	
280						285				290						
ctg	tct	ttg	cat	gtg	ctg	gaa	aca	aat	gat	gaa	aag	gat	caa	aca	gct	1026
Leu	Ser	Leu	His	Val	Leu	Glu	Thr	Asn	Asp	Glu	Lys	Asp	Gln	Thr	Ala	
295						300				305						
aat	ttg	acc	aat	caa	gga	aaa	aat	att	acc	aaa	aat	gtg	act	ggt	ttt	1074
Asn	Leu	Thr	Asn	Gln	Gly	Lys	Asn	Ile	Thr	Lys	Asn	Val	Thr	Gly	Phe	
310						315				320		325				
ttc	cag	tct	ttt	aaa	agc	atc	ctt	act	aat	ccc	ctg	tat	gtt	atg	ttt	1122
Phe	Gln	Ser	Phe	Lys	Ser	Ile	Leu	Thr	Asn	Pro	Leu	Tyr	Val	Met	Phe	
330						335				340						
gtg	ctt	ttg	acg	ttg	tta	caa	gta	agc	agc	tat	att	ggt	gct	ttt	act	1170
Val	Leu	Leu	Thr	Leu	Leu	Gln	Val	Ser	Ser	Tyr	Ile	Gly	Ala	Phe	Thr	
345						350				355						
tat	gtc	ttc	aaa	tac	gta	gag	caa	cag	tat	ggt	cag	cct	tca	tct	aag	1218
Tyr	Val	Phe	Lys	Tyr	Val	Glu	Gln	Gln	Tyr	Gly	Gln	Pro	Ser	Ser	Lys	
360						365				370						
gct	aac	atc	tta	ttg	gga	gtc	ata	acc	ata	cct	att	ttt	gca	agt	gga	1266
Ala	Asn	Ile	Leu	Leu	Gly	Val	Ile	Thr	Ile	Pro	Ile	Phe	Ala	Ser	Gly	
375						380				385						
atg	ttt	tta	gga	gga	tat	atc	att	aaa	aaa	ttc	aaa	ctg	aac	acc	gtt	1314
Met	Phe	Leu	Gly	Gly	Tyr	Ile	Ile	Lys	Lys	Phe	Lys	Leu	Asn	Thr	Val	
390						395				400		405				

gga att gcc aaa ttc tca tgt ttt act gct gtg atg tca ttg tcc ttt Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val Met Ser Leu Ser Phe 410 415 420	1362
tac cta tta tat ttt ttc ata ctc tgt gaa aac aaa tca gtt gcc gga Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn Lys Ser Val Ala Gly 425 430 435	1410
cta acc atg acc tat gat gga aat aat cca gtg aca tct cat aga gat Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val Thr Ser His Arg Asp 440 445 450	1458
gta cca ctt tct tat tgc aac tca gac tgc aat tgt gat gaa agt caa Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn Cys Asp Glu Ser Gln 455 460 465	1506
tgg gaa cca gtc tgt gga aac aat gga ata act tac atc tca ccc tgt Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr Tyr Ile Ser Pro Cys 470 475 480 485	1554
cta gca ggt tgc aaa tct tca agt ggc aat aaa aag cct ata gtg ttt Leu Ala Gly Cys Lys Ser Ser Gly Asn Lys Lys Pro Ile Val Phe 490 495 500	1602
tac aac tgc agt tgt ttg gaa gta act ggt ctc cag aac aga aat tac Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu Gln Asn Arg Asn Tyr 505 510 515	1650
tca gcc cat ttg ggt gaa tgc cca aga gat gat gct tgt aca agg aaa Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp Ala Cys Thr Arg Lys 520 525 530	1698
ttt tac ttt ttt gtt gca ata caa gtc ttg aat tta ttt ttc tct gca Phe Tyr Phe Val Ala Ile Gln Val Leu Asn Leu Phe Phe Ser Ala 535 540 545	1746
ctt gga ggc acc tca cat gtc atg ctg att gtt aaa att gtt caa cct Leu Gly Gly Thr Ser His Val Met Leu Ile Val Lys Ile Val Gln Pro 550 555 560 565	1794
gaa ttg aaa tca ctt gca ctg ggt ttc cac tca atg gtt ata cga gca Glu Leu Lys Ser Leu Ala Leu Gly Phe His Ser Met Val Ile Arg Ala 570 575 580	1842
cta gga gga att cta gct cca ata tat ttt ggg gct ctg att gat aca Leu Gly Gly Ile Leu Ala Pro Ile Tyr Phe Gly Ala Leu Ile Asp Thr 585 590 595	1890
acg tgt ata aag tgg tcc acc aac aac tgt ggc aca cgt ggg tca tgt Thr Cys Ile Lys Trp Ser Thr Asn Asn Cys Gly Thr Arg Gly Ser Cys 600 605 610	1938
agg aca tat aat tcc aca tca ttt tca agg gtc tac ttg ggc ttg tct Arg Thr Tyr Asn Ser Thr Ser Phe Ser Arg Val Tyr Leu Gly Leu Ser 615 620 625	1986
tca atg tta aga gtc tca tca ctt gtt tta tat att ata tta att tat Ser Met Leu Arg Val Ser Ser Leu Val Leu Tyr Ile Ile Leu Ile Tyr 630 635 640 645	2034

gcc atg aag aaa aaa tat caa gag aaa gat atc aat gca tca gaa aat Ala Met Lys Lys Lys Tyr Gln Glu Lys Asp Ile Asn Ala Ser Glu Asn 650 655 660	2082
gga agt gtc atg gat gaa gca aac tta gaa tcc tta aat aaa aat aaa Gly Ser Val Met Asp Glu Ala Asn Leu Glu Ser Leu Asn Lys Asn Lys 665 670 675	2130
cat ttt gtc cct tct gct ggg gca gat agt gaa aca cat tgt taa His Phe Val Pro Ser Ala Gly Ala Asp Ser Glu Thr His Cys * 680 685 690	2175
ggggagaaaa aaagccactt ctgcctctgt gtttccaaac agcattgcat tgattcagta agatgttatt tttgaggagt tcctggctct ttcactaaga atttccacat cttttatggt ggaagtataa ataagcctat gaacttataa taaaacaaac tgttaggtaga aaaaaaaaaa aaaaaa	2235 2295 2355 2361
<p><210> 10 <211> 691 <212> PRT <213> H. sapiens</p> <p><220> <221> VARIANT <222> (130)...(130) <223> Xaa = Asp or Asn</p> <p><221> VARIANT <222> (191)...(191) <223> Xaa = Leu</p> <p><400> 10</p>	
Met Asp Gln Asn Gln His Leu Asn Lys Thr Ala Glu Ala Gln Pro Ser 1 5 10 15 Glu Asn Lys Lys Thr Arg Tyr Cys Asn Gly Leu Lys Met Phe Leu Ala 20 25 30 Ala Leu Ser Leu Ser Phe Ile Ala Lys Thr Leu Gly Ala Ile Ile Met 35 40 45 Lys Ser Ser Ile Ile His Ile Glu Arg Arg Phe Glu Ile Ser Ser Ser 50 55 60 Leu Val Gly Phe Ile Asp Gly Ser Phe Glu Ile Gly Asn Leu Leu Val 65 70 75 80 Ile Val Phe Val Ser Tyr Phe Gly Ser Lys Leu His Arg Pro Lys Leu 85 90 95 Ile Gly Ile Gly Cys Phe Ile Met Gly Ile Gly Gly Val Leu Thr Ala 100 105 110 Leu Pro His Phe Phe Met Gly Tyr Tyr Arg Tyr Ser Lys Glu Thr Asn 115 120 125 Ile Xaa Ser Ser Glu Asn Ser Thr Ser Thr Leu Ser Thr Cys Leu Ile 130 135 140 Asn Gln Ile Leu Ser Leu Asn Arg Ala Ser Pro Glu Ile Val Gly Lys 145 150 155 160 Gly Cys Leu Lys Glu Ser Gly Ser Tyr Met Trp Ile Tyr Val Phe Met 165 170 175 Gly Asn Met Leu Arg Gly Ile Gly Glu Thr Pro Ile Val Pro Xaa Gly 180 185 190 Leu Ser Tyr Ile Asp Asp Phe Ala Lys Glu Gly His Ser Ser Leu Tyr 195 200 205 Leu Gly Ile Leu Asn Ala Ile Ala Met Ile Gly Pro Ile Ile Gly Phe 210 215 220 Thr Leu Gly Ser Leu Phe Ser Lys Met Tyr Val Asp Ile Gly Tyr Val 225 230 235 240	

Asp Leu Ser Thr Ile Arg Ile Thr Pro Thr Asp Ser Arg Trp Val Gly
 245 250 255
 Ala Trp Trp Leu Asn Phe Leu Val Ser Gly Leu Phe Ser Ile Ile Ser
 260 265 270
 Ser Ile Pro Phe Phe Leu Pro Gln Thr Pro Asn Lys Pro Gln Lys
 275 280 285
 Glu Arg Lys Ala Ser Leu Ser Leu His Val Leu Glu Thr Asn Asp Glu
 290 295 300
 Lys Asp Gln Thr Ala Asn Leu Thr Asn Gln Gly Lys Asn Ile Thr Lys
 305 310 315 320
 Asn Val Thr Gly Phe Phe Gln Ser Phe Lys Ser Ile Leu Thr Asn Pro
 325 330 335
 Leu Tyr Val Met Phe Val Leu Leu Thr Leu Leu Gln Val Ser Ser Tyr
 340 345 350
 Ile Gly Ala Phe Thr Tyr Val Phe Lys Tyr Val Glu Gln Gln Tyr Gly
 355 360 365
 Gln Pro Ser Ser Lys Ala Asn Ile Leu Leu Gly Val Ile Thr Ile Pro
 370 375 380
 Ile Phe Ala Ser Gly Met Phe Leu Gly Gly Tyr Ile Ile Lys Lys Phe
 385 390 395 400
 Lys Leu Asn Thr Val Gly Ile Ala Lys Phe Ser Cys Phe Thr Ala Val
 405 410 415
 Met Ser Leu Ser Phe Tyr Leu Leu Tyr Phe Phe Ile Leu Cys Glu Asn
 420 425 430
 Lys Ser Val Ala Gly Leu Thr Met Thr Tyr Asp Gly Asn Asn Pro Val
 435 440 445
 Thr Ser His Arg Asp Val Pro Leu Ser Tyr Cys Asn Ser Asp Cys Asn
 450 455 460
 Cys Asp Glu Ser Gln Trp Glu Pro Val Cys Gly Asn Asn Gly Ile Thr
 465 470 475 480
 Tyr Ile Ser Pro Cys Leu Ala Gly Cys Lys Ser Ser Ser Gly Asn Lys
 485 490 495
 Lys Pro Ile Val Phe Tyr Asn Cys Ser Cys Leu Glu Val Thr Gly Leu
 500 505 510
 Gln Asn Arg Asn Tyr Ser Ala His Leu Gly Glu Cys Pro Arg Asp Asp
 515 520 525
 Ala Cys Thr Arg Lys Phe Tyr Phe Phe Val Ala Ile Gln Val Leu Asn
 530 535 540
 Leu Phe Phe Ser Ala Leu Gly Gly Thr Ser His Val Met Leu Ile Val
 545 550 555 560
 Lys Ile Val Gln Pro Glu Leu Lys Ser Leu Ala Leu Gly Phe His Ser
 565 570 575
 Met Val Ile Arg Ala Leu Gly Gly Ile Leu Ala Pro Ile Tyr Phe Gly
 580 585 590
 Ala Leu Ile Asp Thr Thr Cys Ile Lys Trp Ser Thr Asn Asn Cys Gly
 595 600 605
 Thr Arg Gly Ser Cys Arg Thr Tyr Asn Ser Thr Ser Phe Ser Arg Val
 610 615 620
 Tyr Leu Gly Leu Ser Ser Met Leu Arg Val Ser Ser Leu Val Leu Tyr
 625 630 635 640
 Ile Ile Leu Ile Tyr Ala Met Lys Lys Tyr Gln Glu Lys Asp Ile
 645 650 655
 Asn Ala Ser Glu Asn Gly Ser Val Met Asp Glu Ala Asn Leu Glu Ser
 660 665 670
 Leu Asn Lys Asn Lys His Phe Val Pro Ser Ala Gly Ala Asp Ser Glu
 675 680 685
 Thr His Cys
 690

<210> 11
 <211> 23
 <212> DNA

<213> H. sapiens	
<400> 11	
ggggctctga ttgataacaac gtg	23
<210> 12	
<211> 30	
<212> DNA	
<213> H. sapiens	
<400> 12	
actgtggcac acgtgggtca tgtaggacat	30
<210> 13	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 13	
ctgctgccaa ctaacattgc	20
<210> 14	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 14	
cacacactaa ccatgcctct	20
<210> 15	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 15	
tccagtcat tggctttgcac	20
<210> 16	
<211> 23	
<212> DNA	
<213> H. sapiens	
<400> 16	
aagaaccaat aaagctgctt act	23
<210> 16	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 16	
gtgtttgcta gccacacccat	20
<210> 17	
<211> 20	
<212> DNA	
<213> H. sapiens	
<400> 17	
ggcaacacccat cctcaaaatgt	20
<210> 18	

<211> 20
<212> DNA
<213> H. sapiens

<400> 18
gatgcttcc tctgtgcagt 20

<210> 19
<211> 20
<212> DNA
<213> H. sapiens

<400> 19
ccttcaagcc gaagaaggct 20

<210> 20
<211> 20
<212> DNA
<213> H. sapiens

<400> 20
aggagttcct ggtccttca 20

<210> 21
<211> 20
<212> DNA
<213> H. sapiens

<400> 21
caagctagac ttcaggcctt 20

<210> 22
<211> 24
<212> DNA
<213> H. sapiens

<400> 22
gaggaattct agctccaata tatt 24

<210> 23
<211> 21
<212> DNA
<213> H. sapiens

<400> 23
gtcctacatg acccacgtgt g 21